PATENT APPLICATION
Attorney Docket No. 15966-776 CIP (CURA-276 CIP)

NOVEL HUMAN PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND METHODS OF USING THE SAME

RELATED APPLICATIONS

This application claims priority to USSNs 60/198,293(15966-776), filed April 19, 2000; 60/198,645 (15966-777), filed April 20, 2000; 60/210,809 (15966-778A), filed June 9, 2000; 60/199,476 (15966-778), filed April 26, 2000; 60/200,025 (15966-779), filed April 26, 2000; 60/224,610 (15966-780A), filed August 11, 2000; 60/200,024 (15966-780), filed April 26, 2000; 60/199,880 (15966-781) filed April 26, 2000; 60/218,591(21402-059), filed July 17, 2000; 60/271,814 (21402-059A), filed February 27, 2001; 60/215,855, filed July 3, 2000 (21402-048); and 09/839,446, filed April 19, 2001 (15966-776 Utility). The contents of these applications are incorporated by reference in their entirety.

FIELD OF THE INVENTION

The invention relates to polynucleotides and the polypeptides encoded by such polynucleotides, as well as vectors, host cells, antibodies and recombinant methods for producing the polypeptides and polynucleotides, as well as methods for using the same.

BACKGROUND OF THE INVENTION

The present invention is based in part on nucleic acids encoding proteins that are new members of the following protein families: gamma aminobutyric acid (GABA) receptor, epidermal growth factor (EGF), complement receptor, hematopoietic stem and progenitor cell (HSPC) protein, sulfotransferase (ST), sytaxin and prohibitin.

The GABA receptor family is a related group of ligand-gated chloride channels, where ligand binding results in chloride ion influx and a change in cell polarization. GABA receptors function as the major inhibitory neurotransmitter receptors in the brain, retina and elsewhere in the central nervous system. Alterations in GABA receptors are associated with a number of clinically relevant events and/or pathologies, including *e.g.* stroke, Huntington's disease, Alzheimer's disease, Parkinson's disease, amyotrophic lateral sclerosis, epilepsy, alcoholism, cardiomyopathy, and depression.

The EGF protein family is a wide category of proteins, often receptors, sharing an EGF-domain important in protein-protein and other interactions. EGF-like receptor proteins are often tyrosine kinases involved in cell proliferation, differentiation, and death. Proteins having EGF domains modulate cell shape and motility, and adhesion. Alterations in EGF-like proteins are associated with a number of clinically relevant events and/or pathologies, including *e.g.* cancer, aberrant angiogenesis, renal disease, and diabetes.

The complement receptor family of proteins. Complement receptors are found on the extracellular surface of peripheral white blood cells, *e.g.* neutrophils and eosinophils. Complement receptor proteins are important in cell adhesion and activation. The levels of several complement receptor proteins are elevated on circulating granulocytes in asthmatic individuals.

Members of the HSPC protein family are expressed in stem cells and progenitor cells giving rise to several cell types, including hematopoietic cells. HSPC proteins may be modulated by cyclosporin A via inhibition of gamma interferon production by T cells. HSPC proteins may be involved in the progression of leukemia, lupus, and anemia.

The ST family of proteins are a group of related enzymes that catalyze the sulfate conjugation of many substances, *e.g.* drugs, xenobiotic compounds, hormones and neurotransmitters. ST proteins may share a common structural motif that is important in enzymatic activity. Alterations in ST proteins may be associated with diseases and/or disorders of the liver, intestine and kidney, *e.g.* primary biliary cirrhosis, cholangitis, hepatitis, ulcers, hyperthyroidism, and developmental disorders.

The sytaxin protein family appear to be involved in the docking of cytoplasmic vesicles with the plasma membrane. These proteins may affect synaptic transmission in the brain and other parts of the central nervous system. Sytaxin proteins may be altered in clincally relevant neurological events and/or pathologies such as Lambert-Eaton myasthenic syndrome, asthma, myxoid liposarcoma, acute myeloid leukemia, and diabetes.

The prohibitin family of tumor-suppressor proteins inhibit cell proliferation, and several human cancers, *e.g.* breast, ovarian, liver and lung, show loss of heterozygosity at the

prohibitin gene locus, suggesting mutations in members of the prohibitin family are associated with cancer onset and/or progression.

SUMMARY OF THE INVENTION

The invention is based, in part, upon the discovery of novel nucleic acids and secreted polypeptides encoded thereby. The nucleic acids and polypeptides are collectively referred to herein as "POLYX" nucleic acids and polypeptides.

Accordingly, in one aspect, the invention includes an isolated nucleic acid that encodes a POLYX polypeptide, or a fragment, homolog, analog or derivative thereof. For example, the nucleic acid can encode a polypeptide at least 85% identical to a polypeptide comprising the amino acid sequences of SEQ ID NO:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32 and/or 34. The nucleic acid can be, *e.g.*, a genomic DNA fragment or a cDNA molecule. In some embodiments, the invention provides an isolated nucleic acid molecule that includes the nucleic acid sequence of any of SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31 and/or 33.

Also included within the scope of the invention is a vector containing one or more of the nucleic acids described herein, and a cell containing the vectors or nucleic acids described herein.

The invention is also directed to host cells transformed with a vector comprising any of the nucleic acid molecules described above.

In another aspect, the invention includes a pharmaceutical composition that includes a POLYX nucleic acid and a pharmaceutically acceptable carrier or diluent.

In a further aspect, the invention includes a substantially purified POLYX polypeptide, *e.g.*, any of the POLYX polypeptides encoded by a POLYX nucleic acid, and fragments, homologs, analogs, and derivatives thereof. The invention also includes a pharmaceutical composition that includes a POLYX polypeptide and a pharmaceutically acceptable carrier or diluent.

In a still a further aspect, the invention provides an antibody that binds specifically to a POLYX polypeptide. The antibody can be, *e.g.*, a monoclonal or polyclonal antibody, and fragments, homologs, analogs, and derivatives thereof. The invention also includes a pharmaceutical composition including POLYX antibody and a pharmaceutically acceptable carrier or diluent. The invention is also directed to isolated antibodies that bind to an epitope on a polypeptide encoded by any of the nucleic acid molecules described above.

The invention also includes kits comprising any of the pharmaceutical compositions described above.

The invention further provides a method for producing a POLYX polypeptide by providing a cell containing a POLYX nucleic acid, *e.g.*, a vector that includes a POLYX nucleic acid, and culturing the cell under conditions sufficient to express the POLYX polypeptide encoded by the nucleic acid. The expressed POLYX polypeptide is then recovered from the cell. Preferably, the cell produces little or no endogenous POLYX polypeptide. The cell can be, *e.g.*, a prokaryotic cell or eukaryotic cell.

The invention is also directed to methods of identifying a POLYX polypeptide or nucleic acids in a sample by contacting the sample with a compound that specifically binds to the polypeptide or nucleic acid, and detecting complex formation, if present.

The invention further provides methods of identifying a compound that modulates the activity of a POLYX polypeptide by contacting a POLYX polypeptide with a compound and determining whether the POLYX polypeptide activity is modified.

The invention is also directed to compounds that modulate POLYX polypeptide activity identified by contacting a POLYX polypeptide with the compound and determining whether the compound modifies activity of the POLYX polypeptide, binds to the POLYX polypeptide, or binds to a nucleic acid molecule encoding a POLYX polypeptide.

In a another aspect, the invention provides a method of determining the presence of, or predisposition to a POLYX-associated disorder in a subject. The method includes providing a sample from the subject and measuring the amount of POLYX polypeptide in the subject

sample. The amount of POLYX polypeptide in the subject sample is then compared to the amount of POLYX polypeptide in a control sample. An alteration in the amount of POLYX polypeptide in the subject protein sample relative to the amount of POLYX polypeptide in the control protein sample indicates the subject has a tissue proliferation-associated condition. A control sample is preferably taken from a matched individual, *i.e.*, an individual of similar age, sex, or other general condition but who is not suspected of having a tissue proliferation-associated condition. Alternatively, the control sample may be taken from the subject at a time when the subject is not suspected of having a tissue proliferation-associated disorder. In some embodiments, the POLYX is detected using a POLYX antibody.

In a further aspect, the invention provides a method of determining the presence of, or predisposition to, a POLYX-associated disorder in a subject. The method includes providing a nucleic acid sample (*e.g.*, RNA or DNA, or both) from the subject and measuring the amount of the POLYX nucleic acid in the subject nucleic acid sample. The amount of POLYX nucleic acid sample in the subject nucleic acid is then compared to the amount of POLYX nucleic acid in a control sample. An alteration in the amount of POLYX nucleic acid in the sample relative to the amount of POLYX in the control sample indicates the subject has a tissue proliferation-associated disorder.

In a still further aspect, the invention provides a method of treating or preventing or delaying a POLYX-associated disorder. The method includes administering to a subject in which such treatment or prevention or delay is desired a POLYX nucleic acid, a POLYX polypeptide, or a POLYX antibody in an amount sufficient to treat, prevent, or delay a tissue proliferation-associated disorder in the subject.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the invention, suitable methods and materials are described below. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In the case of conflict, the present

Specification, including definitions, will control. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

Other features and advantages of the invention will be apparent from the following detailed description and claims.

DETAILED DESCRIPTION OF THE INVENTION

The invention provides novel polynucleotides and the polypeptides encoded thereby. The invention is based in part on the discovery of nucleic acids encoding 17 proteins that are novel members of the following protein families: gamma aminobutyric acid (GABA) receptor, epidermal growth factor (EGF), complement receptor, hematopoeitic stem and progenitor cell (HSPC), sulfotransferase, syntaxin, and prohibitin. These nucleic acids, and their associated polypeptides, antibodies and other compositions are referred to as POLY1, POLY2, POLY3 through POLY17, respectively. These sequences are collectively referred to as "POLYX nucleic acids" or "POLYX polynucleotides" (where X is an integer between 1 and 17) and the corresponding encoded polypeptide is referred to as a "POLYX polypeptide" or "POLYX protein".

POLY1-4 are novel members of the GABA receptor family; POLY5-8 are novel members of the EGF family; POLY9-11 are novel members of the complement receptor family; POLY12 is a novel member of the HSPC family; POLY13 is a novel member of the sulfotransferase family; POLY14-16 are novel members of the syntaxin family; and POLY17 is a novel member of the prohibitin family.

Table 1 provides a cross-reference between a POLYX nucleic acid or polypeptide of the invention, a table disclosing a nucleic acid and encoded polypeptide that is encompassed by an indicated POLYX nucleic acid or polypeptide of the invention, and a corresponding sequence identification number (SEQ ID NO:). Also provided is a CuraGen internal Clone Identification Number for the disclosed nucleic acid and encoded polypeptides. Unless indicated otherwise, reference to a "Clone" herein refers to a discrete *in silico* nucleic acid sequence.

TABLE 1.

Clone	POLYX Number	Table Number	SEQ ID NO: Nucleic Acid	SEQ ID NO: Polypeptide
GM_83055392_A	1	2	1	2
83055392	2	3	3	4
CG54683-02	3	4	5	6
CG54683-03	4	5	7	8
Z97832_B.0.704	5	6	9	10
Z97832_B.0.707	6	7	11	12
Z97832_B_1	7	8	13	14
CG55096-04	8	9	15	16
10327789.0.16	9	10	17	18
10327789.0.140	10	11	19	20
10327789_1	11	12	21	22
AC016030_A.0.82	12	13	23	24
h_nh0443k08_A	13	14	25	26
h_nh0778p17_A	14	15	27	28
hnh0778p17_A1	15	16	29	30
CG55655_02	16	17	31	32
GM_11817402_A	17	18	33	34

POLYX nucleic acids, POLYX polypeptides, POLYX antibodies, and related compounds, are useful in a variety of applications and contexts. For example, various POLYX nucleic acids and polypeptides according to the invention are useful, *inter alia*, as novel members of the protein families according to the presence of domains and sequence relatedness to previously described proteins.

POLYX nucleic acids and polypeptides according to the invention can also be used to identify cell types based on the presence or absence of various POLYX nucleic acids according to the invention. Additional utilities for POLYX nucleic acids and polypeptides are discussed below.

POLY1-POLY4

GAMMA AMINOBUTYRIC ACID RECEPTOR-LIKE (GABA) NUCLEIC ACIDS AND PROTEINS

POLY1-4 nucleic acids and proteins are members of the gamma-aminobutyric acid receptor family. GABA receptors are a family of ligand-gated chloride channels that are the major inhibitory neurotransmitter receptors in the nervous system. Gamma-aminobutyric acid

(GABA) is the major inhibitory neurotransmitter of the brain and acts through binding to GABA(A) receptors, where the ligand causes an influx of chloride ions. Certain GABA receptor sub-types are inhibitory neurotransmitter receptors of the brain, the retina, or other parts of the CNS.

GABA receptors are molecular substrates for the regulation of vigilance, anxiety, muscle tension, epileptogenic activity, and memory functions. This is evident since GABA receptors are the site of action of a number of important pharmacological agents, including barbiturates, benzodiazepines, and ethanol. Accordingly, benzodiazepino-induced behavioral responses are mediated by specific GABA(A) receptor sub-types in distinct neuronal circuits.

GABA(A) receptors are heterooligomeric, and combinations of different subunits lead to functional diversity. The gene encoding the gamma-3 form of the GABA receptor (GABRG3) is located on 15q11-q13 in a cluster with GABRA5 and GABRB3. Thus, there is an alpha/beta/gamma cluster of GABA receptor subunit subtype genes on 3 chromosomes, 15, 5, and 3. It has been suggested that these may have originated from chromosome 15 because the centromere of that chromosome is associated with increased amounts of satellite DNA and translocations occur more frequently around such centromeres. Thus, it can be speculated that the ancestral GABA-A receptor gene cluster formed on chromosome 15 and thereafter spawned the clusters on chromosomes 4 and 5. The detailed physical map of this GABAA receptor subunit gene cluster should not only be useful in genetic studies of the 15q11-q13 region, but will also be important for investigating the evolution and expression of the GABAA receptor gene superfamily. Information regarding the known properties of the GABA receptor family, to which POLY1-4 belong, is described in detail below.

<u>Retinal Inhibitory Receptor Properties of GABA receptors</u>

Certain GABA receptor sub-types have been shown to be inhibitory receptors in the retina. Gamma-aminobutyrate is the gamma-aminobutyric acid (GABA) receptor subunit. GABArho1delta51 is an alternatively spliced form of the GABArho1 receptor that was recently isolated from human retina cDNA libraries. The rho1delta51 receptor subunit lacks 17 amino acids in the extracellular N-terminal domain and, when expressed in Xenopus oocytes, forms functional homomeric GABA receptors. Unexpectedly, even after a such a big deletion,

the fundamental properties of the deleted variant receptors are very similar to those of the complete GABArho1 receptors. For example, both types of receptors are bicuculline resistant, desensitize very little, and are negatively modulated by Zn2+ and positively modulated by La3+. In spite of such similarities, the GABArho1delta51 receptors are more sensitive to GABA, to the specific GABA(C) antagonist (1,2,5,6-tetrahydropyridine-4-yl)methylphosphinic acid and to Zn2+, than the complete GABArho1 receptors. The GABArho1delta51 receptors extend the variety of inhibitory receptors in the retina.

GABA receptor rho3 subunit has been localized to rat retina and has been shown to be expressed in rat retina. Digoxigenin-labelled single strand DNA probes was used to examine the expression of the mRNA encoding gamma-aminobutyric acid receptor rho 3 subunit in sections of the adult rat retina. Transcript for the rho 3 subunit was found in cell somata of a portion of cells lying in the ganglion cell layer.

Additionally, cloned cDNA encoding a putative member of GABA receptor rho-subunit class was isolated from rat-retina-mRNA-derived libraries. The cDNA encodes a signal peptide of 21 amino acids followed by the mature rho 3 subunit sequence of 443 amino acids. The proposed amino acid sequence exhibits 63 and 61% homology to the previously-reported human rho 1 and rat rho 2 sequences, respectively. Northern blot analysis demonstrated the expression of mRNA for rho 3 subunit in retina.

Type A gamma-aminobutyric acid (GABA-A) receptors are a family of ligand- gated chloride channels that are the major inhibitory neurotransmitter receptors in the nervous system. Molecular cloning has revealed diversity in the subunits that compose this heterooligomeric receptor, but each previously elucidated subunit displays amino acid similarity in conserved structural elements. These highly conserved regions were used to identify additional members of this family by using the polymerase chain reaction (PCR). One PCR product was used to isolate a full-length cDNA from a human retina cDNA library. The mature protein predicted from this cDNA sequence is 458 amino acids long and displays between 30 and 38% amino acid similarity to the previously identified GABAA subunits. This gene is expressed primarily in the retina but transcripts are also detected in the brain, lung, and thymus. Injection of Xenopus oocytes with RNA transcribed in vitro produces a GABA-

responsive chloride conductance and expression of the cDNA in COS cells yields GABA-displaceable muscimol binding. These features are consistent with the identification of a GABA subunit, GABA rho 1, with prominent retinal expression that increases the diversity and tissue specificity of this ligand-gated ion-channel receptor family.

Intracellular Trafficking of GABA(A) Receptors

Some of the mechanisms that control the intracellular trafficking of GABA(A) receptors have recently been described. Following the synthesis of alpha, beta, and gamma subunits in the endoplasmic reticulum, ternary receptor complexes assemble slowly and are inefficiently inserted into surface membranes of heterologous cells. While beta3, beta4, and gamma2S subunits appear to contain polypeptide sequences that alone are sufficient for surface targeting, these sequences are neither conserved nor essential for surface expression of heteromeric GABA(A) receptors formed from alpha1beta or alpha1betagamma subunits. At the neuronal surface, native GABA(A) receptor clustering and synaptic targeting require a gamma2 subunit and the participation of gephyrin, a clustering protein for glycine receptors. A linker protein, such as the GABA(A) receptor associated protein (GABARAP), may be necessary for the formation of GABA(A) receptor aggregates containing gephyrin. A substantial fraction of surface receptors are sequestered by endocytosis, another process which requires a GABA(A) receptor gamma2 subunit. In heterologous cells, constitutive endocytosis seems to predominate while, in cortical neurons, internalization is evoked when receptors are occupied by GABA(A) agonists. After constitutive endocytosis, receptors are relatively stable and can be rapidly recycled to the cell surface, a process that may be regulated by protein kinase C. On the other hand, a portion of the intracellular GABA(A) receptors derived from ligand-dependent endocytosis is apparently degraded. The clustering of GABA(A) receptors at synapses and at coated pits are two mechanisms that may compete for a pool of diffusable receptors, providing a model for plasticity at inhibitory synapses.

Chromosomal Localizatin of GABA Receptors

The gamma-aminobutyric acid (GABA) receptors are the major inhibitory neurotransmitter receptors in the brain and the site of action of a number of important pharmacological agents including barbiturates, benzodiazepines, and ethanol. The gamma 1

and gamma 2 subunits have been shown to be important in mediating responses to benzodiazepines, and a splicing variant of the gamma 2 subunit, gamma 2L, has been shown to be necessary for ethanol actions on the receptor, raising the possibility that the gamma 2 gene may be involved in human genetic predisposition to the development of alcoholism. The human genes encoding the gamma 1 and gamma 2 subunits of the GABAA receptor has been mapped to chromosomes 4 and 5, respectively, by PCR amplification of human-specific products from human-hamster somatic cell hybrid DNAs. Using panels of chromosome-specific natural deletion hybrids, the gamma 1 gene (GABRG1) has been further localized to 4p14-q21.1 and the gamma 2 gene (GABRG2) to 5q31.1-q33.2. This localization indicates that the gamma 1 gene may be clustered together with the previously mapped alpha 2 and beta 1 genes on chromosome 4 and that the gamma 2 gene may be close to the previously localized alpha 1 gene on chromosome 5. To further examine the latter possibility, the alpha 1 gene was mapped using the chromosome 5 deletion hybrids, and was shown to be within the same region as the gamma 2 gene, 5q31.1-q33.2. A PCR-based screening strategy was used to isolate a 450-kilobase human genomic yeast artificial chromosome clone containing both the alpha 1 and gamma 2 genes. Pulsed-field gel restriction mapping of the yeast artificial chromosome indicated that the two genes are within 200 kilobases of each other. This demonstrates that members of the GABAA receptor gene family often occur in small gene clusters widely distributed in the genome.

Additionally, genes encoding rho2 (GABRR2) and rho1 (GABRR1) have been localized to human chromosome 6q14-q21 and mouse chromosome 4. Two distinct clones have been identified by screening a genomic DNA library with a portion of the cDNA encoding the gamma-aminobutyric acid (GABA) receptor subunit rho1. DNA sequencing revealed that one clone contained a single exon from the rho1 gene (GABBR1) while the second clone encompassed an exon with 96% identity to the rho1 gene. Screening of a human retina cDNA library with oligonucleotides specific for the exon in the second clone identified a 3-kb cDNA with an open reading frame of 1395 bp. The predicted amino acid sequence of this cDNA demonstrated 30 to 38% similarity to alpha, beta, gamma, and delta GABA receptor subunits and 74% similarity to the GABA rho1 subunit, suggesting that the newly isolated cDNA encoded a new member of the rho subunit family, tentatively named GABA

rho2. Polymerase chain reaction (PCR) amplification of rho1 and rho2 gene sequences from DNA of three somatic cell hybrid panels mapped both genes to human chromosome 6, bands q14 to q21. Tight linkage was also demonstrated between restriction fragment length variants (RFLVs) from each rho gene and the Tsha locus on mouse chromosome 4, which is homologous to the CGA locus on human chromosome 6q12-q21. These two lines of evidence confirmed that GABRR1 and newly identified GABRR2 mapped to the same region on human chromosome 6. This close physical association and high degree of sequence similarity raised the possibility that one rho gene arose from the other by duplication.

Benzodiazepine Actions Mediated by GABA Receptors

When gamma-aminobutyric acid (GABA), the major inhibitory neurotransmitter in vertebrate brain, binds to its receptor, it activates a chloride channel. Neurotransmitter action at the GABAA receptor is potentiated by both benzodiazepines and barbiturates which are therapeutically useful drugs. GABA(A) receptors are therefore molecular substrates for the regulation of vigilance, anxiety, muscle tension, epileptogenic activity and memory functions, which is evident from the spectrum of actions elicited by clinically effective drugs acting at their modulatory benzodiazepine-binding site.

There is strong evidence that this receptor is heterogeneous. Complementary DNAs encoding an alpha- and a beta-subunit have previously isolated. It has been shown that both are needed for expression of a functional GABAA receptor. cDNAs encoding two additional GABAA receptor alpha-subunits have now been isolated, confirming the heterogeneous nature of the receptor/chloride channel complex and demonstrating a molecular basis for it. These alpha-subunits are differentially expressed within the CNS and produce, when expressed with the beta-subunit in Xenopus oocytes, receptor subtypes which can be distinguished by their apparent sensitivity to GABA. Highly homologous receptor subtypes which differ functionally are common feature of brain receptors.

Additionally, evidence shows that benzodiazepine actions are mediated by specific GABA receptor sub-types. By introducing a histidine-to-arginine point mutation at position 101 of the murine alpha1-subunit gene, it has been shown that alpha1-type GABA(A) receptors, which are mainly expressed in cortical areas and thalamus, are rendered insensitive

to allosteric modulation by benzodiazepine-site ligands, while regulation by the physiological neurotransmitter gamma-aminobutyric acid is preserved. Alpha1(H101R) mice failed to show the sedative, amnesic and partly the anticonvulsant action of diazepam. In contrast, the anxiolytic-like, myorelaxant, motor-impairing and ethanol-potentiating effects were fully retained, and are attributed to the nonmutated GABA(A) receptors found in the limbic system (alpha2, alpha5), in monoaminergic neurons (alpha3) and in motoneurons (alpha2, alpha5). Thus, benzodiazepine-induced behavioural responses are mediated by specific GABA(A) receptor subtypes in distinct neuronal circuits, which is of interest for drug design.

Benzodiazepines have come under scrutiny and attack over recent years because of their abuse liability, withdrawal reactions and development of tolerance. Consequently, practitioners worldwide are discouraged from prescribing them. While some of these risks may have been exaggerated, benzodiazepines remain a useful therapeutic tool, alone or in combination, in a number of psychiatric and medical conditions. Withholding such treatment may be unjustified and detrimental to the patients' health. Further, benzodiazepines have helped researchers in their attempts to elucidate the neurobiological mechanisms underlying anxiety. This, in return, leads to the development of new effective anxiolytic treatments, with fewer problems compared to the traditional benzodiazepine compounds. Such new agents are already available or at the closing stages of clinical trials.

Role of GABA Receptor in Cognitive Function

The role of GABA in cognitive functions was also studied in cats which had received damage to the forebrain basal nuclei. The cognitive functions were studied on experimental model of Alzheimer's disease (destruction of the basal nuclei of Meynert in cats) using the stimulation and inhibition of Ach, GABA, and DA brain systems. Ach system was found to be essential to form generalization function, DA system to improve simple learning, and GABA system to involve in formation of complex associations.

Novel members of the GABA receptor family, POLY1-POLY4, are described in detail below. These nucleic acids and proteins function as described above, and therefore are useful in modulating neurological, *e.g.* conditions related to brain neurotransmitters, and other disorders.

The protein similarity information, expression pattern, cellular localization, and map location for POLY1-POLY4 discussed below suggest that these GABA Receptor-like proteins have important structural and/or physiological functions characteristic of the GABA Receptor family. Therefore, the nucleic acids and proteins of the invention are useful in potential diagnostic and therapeutic applications, *e.g.* diagnosis and therapy of neurological diseases and/or disorders, and as research tools. Additionally, POLY1-POLY4 have applications in the diagnosis and/or treatment of various diseases and disorders. For example, the compositions of POLY1-POLY4 will have efficacy for the treatment of patients suffering from: psychiatric and medical conditions, depression, stroke, Parkinson's disease, Huntington's disease, Tourette's syndrome, amyotrophic lateral sclerosis, head trauma, Alzheimer's disease, alcoholism, vigilance, anxiety, muscle tension, epileptogenic activity and memory functions, cardiomyopathy, and arrhythmogenic right ventricular dysplasia as well as other diseases, disorders and conditions.

These materials are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in diagnostic and/or therapeutic methods.

POLY1

A novel nucleic acid was identified on chromosome 3 as described in Example 1. The novel nucleic acid of 1875 nucleotides (SEQ ID NO: 1), which encodes a novel gamma aminobutyric acid receptor -like protein is shown in Table 2A. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 10, 11 and 12 and ending with a TGA codon at nucleotides 1411, 1412 and 1413. A putative untranslated region upstream from the initiation codon and downstream from the termination codon is underlined in Table 2A, and the start and stop codons are in bold letters. The encoded protein having 467 amino acid residues (SEQ ID NO:2) is presented using the one-letter code in Table 2B.

TABLE 2A: The nucleotide sequence of POLY1.

TTGGAAGAG**ATG**GTCCTGGCTTTCCAGTTAGTCTCCTTCACCTACATCTGGATCATATTGAAACCAAATG
TTTGTGCTGCTTCTAACATCAAGATGACACACCAGCGGTGCTCCTCTTCAATGAAACAAACCTGGATGCA
AGAAACTAGAATGAAGAAAGATGACAGTACCAAAGCGCGGCCTCAGAAATATGAGCAACTTCTCCATATA
GAGGACAACGATTTCGCAATGAGACCTGGATTTGGAGGTTCTCCAGTGCCAGTAGGTATAGATGTCCATG
TTGAAAGCATTGACAGCATTTCAGAGACTAACATGGACTTTACAATGACTTTTTATCTCAGGCATTACTG
GAAAGACGAGAGGCTCTCCTTTCCTAGCACAGCAAACAAAAGCATGACATTTGATCATAGACACTTGCGG
TATTCGTTATTCATCAGAAGGCTGTATCTGTTATACTGCCAGAGGTCTTTCTTCTCACCCTCATCCATAC
TTCCCTCATCTCCAGACATCCATGCACCTGGTACATCTAAAAGCAGTTTGTCTGATAGCCTTGTATGTAT
ATCTGAAAAAAACTTGCCAGGACACAGTAAAAACACACCTCTTGCAATGTCAGATGTAGCCTACAATGAG
GATGACCTAATGCTATACTGGAAACACGGAAACAAGTCCTTAAATACTGAAGAACATATGTCCCTTTCTC
AGTTCTTCATTGAAGACTTCAGTGCATCTAGTGGATTAGCTTTCTATAGCAGCACAGGTACAGCATTTTA
CATGGGTGATTCATCAGCATTTATTGGACATCTACTGTTTTTGATCTGGAGTTCCAGGAAAAGACCAGGT
TTAGAGATGTTGGGTTTGGGAATTCTCAGAATCTGGGTAATAACTAGAGCCATGGATAAGAAAATGGAAA
TGGGAATCACCACAGTGCTGACCATGTCCACAATCATCACTGCTGTGAGCGCCTCCATGCCCCAGGTGTC
CTACCTCAAGGCTGTGGATGTGTACCTGTGGGTCAGCTCCCTCTTTGTGTTCCTGTCAGTCATTGAGTAT
GCAGCTGTGAACTACCTCACCACAGTGGAAGAGCGGAAACAATTCAAAAAAAGTTTTTCAAAGATTTCTA
GGATGTACAATATTGATGCAGTTCAAGCTATGGCCTTTGATGGTTGTTACCATGACAGCGAGATTGACAT
GGACCAGACTTCCCTCTCTCTAAACTCAGAAGACTTCATGAGAAGAAAATCGATATGCAGCCCCAGCACC
GATTCATCTCGGATAAAGAGAAGAAAATCCCTAGGAGGACATGTTGGTAGAATCATTCTGGAAAACAACC
ATGTCATTGACACCTATTCTAGGATTTTATTCCCCATTGTGTATATCTTTTTTAATTTGTTTTACTGGGG
TGTATATGTA**TGA**AGGGGAATTTCAAATGTATACAACTTTAAAGCCAGATGATGTTTAAAAACAAAACTC
TTGAATATGAGTTGGATAGTCCTAGATGGAACTGGGAAAGAGCAAGTCACCTCTCCTGCCCTAATGAAAA
TTTGAAAGCTGTCTGATTTACATCTAAGAAAGAGTTTAGGTCCTAGAAAAGTTTGACTCCATAAATAAGA
GTCATAGGCATGTGTATTATGGGAAAAACAGTTTTCCATTGGGAAGGGCTTTATAACTACTTCATCTGAA
CCCTCCTTCTTTCTTAATGAAATGTTCTTTATTTAACTAGGGAAGAAAGCTGGACTATAACAATAATTCA
AAGATATTTTGTTTCTTAGTGCCAGCCAAGTGCCTGGTTATCTACCAGAGCTCAACCGTCCTAGGCAAGA
ACATCCACATAGAGGTGGTATCATCCACACTCACACAGCTGAGAATCCTATGAAG (SEQ ID NO:1)

TABLE 2B: Protein sequence encoded by the coding sequence shown in TABLE 2A

MVLAFQLVSFTYIWIILKPNVCAASNIKMTHQRCSSSMKQTWMQETRMKKDDSTKARPQKYEQLLHIEDNDFAMRPGFGGSPVP
VGIDVHVESIDSISETNMDFTMTFYLRHYWKDERLSFPSTANKSMTFDHRHLRYSLFIRRLYLLYCQRSFFSPSSILPSSPDIH
APGTSKSSLSDSLVCISEKNLPGHSKNTPLAMSDVAYNEDDLMLYWKHGNKSLNTEEHMSLSQFFIEDFSASSGLAFYSSTGTA
FYMGDSSAFIGHLLFLIWSSRKRPGLEMLGLGILRIWVITRAMDKKMEMGITTVLTMSTIITAVSASMPQVSYLKAVDVYLWVS
SLFVFLSVIEYAAVNYLTTVEERKQFKKSFSKISRMYNIDAVQAMAFDGCYHDSEIDMDQTSLSLNSEDFMRRKSICSPSTDSS
RIKRRKSLGGHVGRIILENNHVIDTYSRILFPIVYIFFNLFYWGVYV (SEQ ID NO:2)

Similarities

In a search of sequence databases, it was found, for example, that the nucleic acid sequence (SEQ ID NO:1) has 1030 of 1366 bases (75 %) identical to a *Rattus norvegicus* gamma aminobutyric acid receptor mRNA (GENBANK-ID: D50671). The full amino acid sequence of the protein of the invention was found to have 296 of 471 amino acid residues (62%) identical to, and 349 of 471 residues 73%) positive with, 464 amino acid gamma aminobutyric acid receptor residue protein from *Rattus norvegicus* (ptnr:SWISSPROT-ACC: P50573) (Table 2C).

TABLE 2C: BLASTX of POLY1 against Gamma-Aminobutyric-Acid Receptor RHO-3 Subunit Precursor (GABA(A)Receptor) (SEQ ID NO:35)

```
>gi|1730196|sp|P50573|GAR3_RAT GAMMA-AMINOBUTYRIC-ACID RECEPTOR RHO-3 SUBUNIT
PRECURSOR (GABA(A)
          RECEPTOR)

 Score =  529 bits (1364), Expect = e-149
 Identities = 296/471 (62%), Positives = 349/471 (73%), Gaps = 11/471 (2%)

Query: 1   MVLAFQLVSFTYIWIILKPNVCAASNIKMTHQRCSSSMKQTWMQETRMKKDDSTKARPQK 60
           |||||  |  ||| || |   +  || +|  ||+| || ||| ++||||+||| ||  | |
Sbjct: 1   MVLAFWLAFFTYTWITL---MLDASAVKEPHQQCLSSPKQTRIRETRMRKDDLTKVWPLK 57

Query: 61  YEQLLHIEDNDFAMRPGFGGSPVPVGIDVHVESIDSISETNMDFTMTFYLRHYWKDERLS 120
            ||||||||+||+ |||||||||||||||| ||||||||| ||||||||||||||||||||
Sbjct: 58  REQLLHIEDHDFSTRPGFGGSPVPVGIDVQVESIDSISEVNMDFTMTFYLRHYWKDERLS 117

Query: 121 FPSTANKSMTFDHRHLRYSLFIRRLYLLYCQRXXXXXXXXXXXXXDIHAPGTSKSSL--S 178
           |||| ||||||| | ++  +++  ++ ++ +|               +|  |    ||  +
Sbjct: 118 FPSTTNKSMTFDRRLIQ-KIWVPDIFFVHSKRSFIHDTTVENIMLRVHPDGNVLFSLRIT 176

Query: 179 DSLVCISE-KNLPGHSKNTPLAMSDVAYNEDDLMLYWKHGNKSLNTEEHMSLSQFFIEDF 237
            | +|  +    |  ++|  | +   ||||+|||||||||||||||||+|||||||+|
Sbjct: 177 VSAMCFMDFSRFPLDTQNCSLELESYAYNEEDLMLYWKHGNKSLNTEEHISLSQFFIEEF 236

Query: 238 SASSGLAFYSSTGTAFYMGDSSAFIGHLLFLIWSSRKRPGLEMLGLGILRIWVITRAMDK 297
           |||||||||||  + +  +     |+ | +  +   | + |+ |  +  |+  ||+
Sbjct: 237 SASSGLAFYSSTGWYYRLFINFVLRRHIFFFVLQTY-FPAMLMVMLSWVSFWIDRRAVPA 295

Query: 298 KMEMGITTVLTMSTIITAVSASMPQVSYLKAVDVYLWVSSLFVFLSVIEYAAVNYLTTVE 357
           ++ +|||||||||||+| ||||||||||+||||||+||||||||||||||||||||||||
Sbjct: 296 RVSLGITTVLTMSTIVTGVSASMPQVSYVKAVDVYMWVSSLFVFLSVIEYAAVNYLTTVE 355

Query: 358 ERKQFKKSFSKISRMYNIDAVQAMAFDGCYHDSEIDMDQTSLSLNS-EDFMRRKSICSPS 416
           | ||  +   ||| |||||||||||||||||| | |+||||  |+| || || |   ||
Sbjct: 356 EWKQLNRR-GKISGMYNIDAVQAMAFDGCYHDGETDVDQTSFFLHSEEDSMRTKFTGSPC 414

Query: 417 TDSSRIKRRKSLGGHVGRIILENNHVIDTYSRILFPIVYIFFNLFYWGVYV 467
           |||+|| |||||+|||||||||||||||||+||+||| |||||||+||
Sbjct: 415 ADSSQIK-RKSLGGNVGRIILENNHVIDTYSRIVFPVVYIIFNLFYWGIYV 464 (SEQ ID NO.: 35)
```

In a search of sequence databases, it was also found, for example, that the nucleic acid sequence (SEQ ID NO:1) has 776 of 781 bases (99%) identical to a CuraGen human assembly (s3aq:83055392). The strong (99%) homology between the gene of current invention with the above CuraGen proprietary human assembly strongly suggests that the current invention represents an expressed gene. In a search of sequence databases, it was also found, for example, that the nucleic acid sequence has 623 of 934 bases (66%) identical to a *Homo sapiens* gamma aminobutyric acid receptor mRNA (GENBANK-ID: M62400). The global sequence homology, as defined by the GAP global sequence alignment program using the full length sequence of the best BlastX match and the full length sequence of the protein of the invention, is 64.6% amino acid identity and 70.2% amino acid homology. In addition, this protein contains the following protein domain, as defined by Interpro, at the indicated amino acid positions: Neurotransmitter-gated ion-channel family domain (IPR001175) at amino acid positions 58-134, 301-361, and 441-463.

It was also found that POLY1 had homology to other amino acid sequences shown in the BLASTX data in Table 2D.

Table 2D. BLASTX alignments of POLY1

Sequences producing High-scoring Segment Pairs:	Reading Frame	High Score	Smallest Sum Prob. P(N)	N
R31188 GABA-A receptor beta-2 subunit - Homo sapi...	+1	396	3.7e-43	2
R59866 Human GABA receptor beta2 subunit - Homo s...	+1	396	3.7e-43	2
R93118 Human GABA-A receptor epsilon subunit - Ho...	+1	412	2.0e-37	1
W26464 Human GABA-A receptor epsilon subunit - Ho...	+1	412	2.0e-37	1
B00174 Breast cancer protein BCR3 - Homo sapiens,...	+1	412	2.0e-37	1
W81634 GABA-gated chloride channel TBW-a2 - Helio...	+1	344	5.6e-37	2

High Score is a ranking of homologous polypeptides, and Smallest Sum Prob. Is the likelyhood that the calculated homology occurred by chance.

It was also found that POLY1 had homology to other amino acid sequences shown in the BLASTP data in Table 2E.

Table 2E. BLASTP results for POLY1

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect

GENBANK Acc: AF101037	gamma-aminobutyric-acid receptor rho-3 subunit precursor [Morone americana]	470	210/479 (43%)	302/479 (62%)	1e-93
GENBANK Acc: NP_058987.1	gamma-aminobutyric acid (GABA-A) receptor, subunit rho 1 [Rattus norvegicus]	474	189/421 (44%)	259/421 (60%)	4e-88
GENBANK Acc:NP_032101.1	gamma-aminobutyric acid (GABA-A) receptor, subunit rho 1 [Mus musculus]	474	189/421 (44%)	259/421 (60%)	8e-88
GENBANK Acc: NP_002033.1	gamma-aminobutyric acid (GABA) receptor, rho 1 precursor; [Homo sapiens]	473	187/421 (44%)	258/421 (60%)	5e-87

In a search of CuraGen's human expressed sequence assembly database, assembly(ies) 83055392 (781 nucleotides) was/were identified as having >95% homology to this nucleic acid sequence. This database is composed of the expressed sequences (as derived from isolated mRNA) from more than 96 different tissues. The mRNA is converted to cDNA and then sequenced. These expressed DNA sequences are then pooled in a database and those exhibiting a defined level of homology are combined into a single assembly with a common consensus sequence. The consensus sequence is representative of all member components. Since the nucleic acid of the described invention has >95% sequence identity with the CuraGen assembly, the nucleic acid of the invention represents an expressed gene sequence. This DNA assembly has 1 component and was found by CuraGen to be expressed in fetal brain.

PSORT analysis predicts the protein of the invention to be localized in plasma membrane with a certainty of 0. 64. Using the SIGNALP analysis, it is predicted that the protein of the invention has a signal peptide with most likely cleavage site between pos. 24 and 25 of SEQ ID NO.: 2.

POLY2

A novel nucleic acid was identified on chromosome 3. The novel nucleic acid of 1417 nucleotides (SEQ ID NO:3) encodes a novel gamma aminobutyric acid receptor-like protein that is shown in TABLE 3A. An open reading frame was identified beginning with an ATG

initiation codon at nucleotides 8-10 and ending with a TGA codon at nucleotides 1400-1402. A putative untranslated region upstream from the initiation codon and downstream from the termination codon is underlined in TABLE 3A, and the start and stop codons are in bold letters. The encoded protein having 464 (SEQ ID NO:4) amino acid residues is presented using the one-letter code in TABLE 3B.

TABLE 3A: Nucleotide sequence of POLY2.

GGAAGAG**ATG**GTCCTGGCTTTCCAGTTAGTCTCCTTCACCTACATCTGGATCATATTGGTTTGTGCTGCTTCTAAC
ATCAAGATGACACACCAGCGGTGCTCCTCTTCAATGAAACAAACCAGCAAACAAGAAACTAGAATGAAGAAAGATG
ACAGTACCAAAGCGCGGCCTCAGAAATATGAGCAACTTCTCCATATAGAGGACAACGATTTCGCAATGAGACCTGG
ATTTGGAGGTTCTCCAGTGCCAGTAGGTATAGATGTCCATGTTGAAAGCATTGACAGCATTTCAGAGACTAACATG
GACTTTACAATGACTTTTTATCTCAGGCATTACTGGAAAGACGAGAGGCTCTCCTTTCCTAGCACAGCAAACAAAA
GCATGACATTTGATCATAGATTGACCAGAAAGATCTGGGTGCCTGATATCTTTTTTGTCCACTCTAAAAGATCCTT
CATCCATGATACAACTATGGAGAATATCATGCTGCGCGTACACCCTGATGGAAACGTCCTCCTAAGTCTCAGGATA
ACGGTTTCGGCCATGTGCTTTATGGATTTCAGCAGGTTTCCTCTTGACACTCAAAATTGTTCTCTTGAACTGGAAA
GCGCCTACAATGAGGATGACCTAATGCTATACTGGAAACACGGAAACAAGTCCTTAAATACTGAAGAACATATGTC
CCTTTCTCAGTTCTTCATTGAAGACTTCAGTGCATCTAGTGGATTAGCTTTCTATAGCAGCACAACAGGCTGGTAC
AATAGGCTTTTCATCATCTCTGTGCTAAGGAGGCATGTTTTCTTCTTTGTGCTGCCAACCTATTTCCCAGCCATAT
TGATGGTGATGCTTTCATGGGTTTCATTTTGGATTGACCGAAGAGCTGTTCCTGCAAGAGTTTCCCTGGGAATCAC
CACAGTGCTGACCATGTCCACAATCATCACTGCTGTGAGCGCCTCCATGCCCCAGGTGTCCTACCTCAAGGCTGTG
GATGTGTACCTGTGGGTCAGCTCCCTCTTTGTGTTCCTGTCAGTCATTGAGTATGCAGCTGTGAACTACCTCACCA
CAGTGGAAGAGCGGAAACAATTCAAGAAGACAGGAAAGATTTCTAGGATGTACAATATTGATGCAGTTCAAGCTAT
GGCCTTTGATGGTTGTTACCATGACAGCGAGATTGACATGGACCAGACTTCCCTCTCTCTAAACTCAGAAGACTTC
ATGAGAAGAAAATCGATATGCAGCCCCAGCACCGATTCATCTCGGATAAAGAGAAGAAAATCCCTAGGAGGACATG
TTGGTAGAATCATTCTGGAAAACAACCATGTCATTGACACCTATTCTAGGATTTTATTCCCCATTGTGTATATTTT
ATTTAATTTGTTTTACTGGGGTGTATATGTAT**GA**AGGGGAATTTCAAAT (SEQ ID NO:3)

TABLE 3B: Protein sequence encoded by the coding sequence shown in TABLE 3A

MVLAFQLVSFTYIWIILVCAASNIKMTHQRCSSSMKQTSKQETRMKKDDSTKARPQKYEQLLHIEDNDFAMRPGFG
GSPVPVGIDVHVESIDSISETNMDFTMTFYLRHYWKDERLSFPSTANKSMTFDHRLTRKIWVPDIFFVHSKRSFIH
DTTMENIMLRVHPDGNVLLSLRITVSAMCFMDFSRFPLDTQNCSLELESAYNEDDLMLYWKHGNKSLNTEEHMSLS
QFFIEDFSASSGLAFYSSTTGWYNRLFIISVLRRHVFFFVLPTYFPAILMVMLSWVSFWIDRRAVPARVSLGITTV
LTMSTIITAVSASMPQVSYLKAVDVYLWVSSLFVFLSVIEYAAVNYLTTVEERKQFKKTGKISRMYNIDAVQAMAF
DGCYHDSEIDMDQTSLSLNSEDFMRRKSICSPSTDSSRIKRRKSLGGHVGRIILENNHVIDTYSRILFPIVYILFN
LFYWGVYV (SEQ ID NO:4)

In a search of sequence databases, it was found, for example, that the nucleic acid sequence has 1208 of 1412 bases (85 %) identical to a *Rattus norvegicus* gamma aminobutyric acid receptor mRNA (GENBANK-ID: D50671). The full amino acid sequence of the protein of the invention was found to have 389 of 464 amino acid residues (83%) identical to, and 415 of 464 residues (89 %) positive with, the 464 amino acid residue protein from *Rattus norvegicus* (ptnr:SWISSPROT-ACC: P50573) (Table 3C).

TABLE 3C: BLASTX of POLY2 against Gamma-Aminobutyric-Acid Receptor RHO-3 Subunit Precursor (GABA(A)Receptor) (SEQ ID NO:36)

```
>ptnr:SWISSPROT-ACC:P50573 GAMMA-AMINOBUTYRIC-ACID RECEPTOR RHO-3 SUBUNIT PRECURSOR
(GABA(A) RECEPTOR) - Rattus norvegicus (Rat), 464 aa.

  Plus Strand HSPs:

 Score = 1992 (701.2 bits), Expect = 4.2e-205, P = 4.2e-205
 Identities = 389/464 (83%), Positives = 415/464 (89%), Frame = +2

Query:     8 MVLAFQLVSFTYIWIILVCAASNIKMTHQRCSSSMKQTSKQETRMKKDDSTKARPQKYEQ 187
             |||||  |  ||| ||  |+  || +|  ||+| || |||  +||||+||| ||  | | ||
Sbjct:     1 MVLAFWLAFFTYTWITLMLDASAVKEPHQQCLSSPKQTRIRETRMRKDDLTKVWPLKREQ 60

Query:   188 LLHIEDNDFAMRPGFGGSPVPVGIDVHVESIDSISETNMDFTMTFYLRHYWKDERLSFPS 367
             ||||||+||+ |||||||||||||| |||||||| |||||||||||||||||||||||||
Sbjct:    61 LLHIEDHDFSTRPGFGGSPVPVGIDVQVESIDSISEVNMDFTMTFYLRHYWKDERLSFPS 120

Query:   368 TANKSMTFDHRLTRKIWVPDIFFVHSKRSFIHDTTMENIMLRVHPDGNVLLSLRITVSAM 547
             | ||||||| || +|||||||||||||||||||||+|||||||||||||| |||||||||
Sbjct:   121 TTNKSMTFDRRLIQKIWVPDIFFVHSKRSFIHDTTVENIMLRVHPDGNVLFSLRITVSAM 180

Query:   548 CFMDFSRFPLDTQNCSLELES-AYNEDDLMLYWKHGNKSLNTEEHMSLSQFFIEDFSASS 724
             |||||||||||||||||||| ||||+|||||||||||||||||+|||||||+|||||
Sbjct:   181 CFMDFSRFPLDTQNCSLELESYAYNEEDLMLYWKHGNKSLNTEEHISLSQFFIEEFSASS 240

Query:   725 GLAFYSSTTGWYNRLFIISVLRRHVFFFVLPTYFPAILMVMLSWVSFWIDRRAVPARVSL 904
             |||||||| ||| ||||  |||||+|||||  |||||+|||||||||||||||||||||||
Sbjct:   241 GLAFYSST-GWYYRLFINFVLRRHIFFFVLQTYFPAMLMVMLSWVSFWIDRRAVPARVSL 299

Query:   905 GITTVLTMSTIITAVSASMPQVSYLKAVDVYLWVSSLFVFLSVIEYAAVNYLTTVEERKQ 1084
             ||||||||||+| ||||||||||+||||||+|||||||||||||||||||||||||| ||
Sbjct:   300 GITTVLTMSTIVTGVSASMPQVSYVKAVDVYMWVSSLFVFLSVIEYAAVNYLTTVEEWKQ 359

Query:  1085 FKKTGKISRMYNIDAVQAMAFDGCYHDSEIDMDQTSLSLNSE-DFMRRKSICSPSTDSSR 1261
               + |||| |||||||||||||||| | |+|||| |+|| | || |   ||  |||+
Sbjct:   360 LNRRGKISGMYNIDAVQAMAFDGCYHDGETDVDQTSFFLHSEEDSMRTKFTGSPCADSSQ 419

Query:  1262 IKRRKSLGGHVGRIILENNHVIDTYSRILFPIVYILFNLFYWGVYV 1399
             ||| |||||+|||||||||||||||||+||+|||+|||||||+||
Sbjct:   420 IKR-KSLGGNVGRIILENNHVIDTYSRIVFPVVYIIFNLFYWGIYV 464
```

(SEQ ID NO:36)

In a search of sequence databases, it was also found, for example, that the nucleic acid sequence has 339 of 340 bases (99%) identical to a CuraGen Corporation human assembly (s3aq:83055392). The strong (99%) homology between the gene of current invention with the above CuraGen proprietary human assembly strongly suggests that the current invention represents an expressed gene. In a search of sequence databases, it was also found, for example, that the nucleic acid sequence has 888 of 1273 bases (69%) identical to a *Homo sapiens* gamma aminobutyric acid receptor mRNA (GENBANK-ID: M62400. The global sequence homology, as defined by the GAP global sequence alignment program using the full length sequence of the best BlastX match and the full length sequence of the protein of the invention, is 84% amino acid identity and 86% amino acid homology. In addition, this protein contains the following protein domain, as defined by Interpro, at the indicated amino acid positions: Neurotransmitter-gated ion-channel family domain (IPR001175) at amino acid positions 57 to 361 and 440 to 462.

In a search of CuraGen's human expressed sequence assembly database, assembly(ies) 83055392 (781 nucleotides) was/were identified as having >95% homology to this predicted gene sequence. The procedure is a differential expression and sequencing procedure that normalizes mRNA species in a sample, and is disclosed in U. S. Ser. No. 09/417,386, filed Oct. 13, 1999, incorporated herein by reference in its entirety. This database is composed of the expressed sequences (as derived from isolated mRNA) from more than 96 different tissues. The mRNA is converted to cDNA and then sequenced. These expressed DNA sequences are then pooled in a database and those exhibiting a defined level of homology are combined into a single assembly with a common consensus sequence.

PSORT analysis predicts the protein of the invention to be localized in plasma membrane with a certainty of 0.68. Using the SIGNALP analysis, it is predicted that the protein of the invention has a signal peptide with most likely cleavage site between residues 22 and 23 of SEQ ID NO. 4.

POLY3

A POLY3 nucleic acid was cloned as described in Example 2. The novel nucleic acid of 1444 nucleotides (SEQ ID NO:5) encodes a novel GABA Receptor-like protein that is

shown in TABLE 4. An open reading frame was identified beginning at nucleotides 21-23 and ending at nucleotides 1425-1427. This polypeptide represents a novel functional GABA Receptor-like protein. The start and stop codons of the open reading frame are highlighted in bold type. Putative untranslated regions (underlined), are found upstream from the initiation codon and downstream from the termination codon. The encoded protein having 468 amino acid residues (SEQ ID NO:6) is presented using the one-letter code in TABLE 4B. Single nucleotide polymorphisms of a POLY3 nucleic acid are described in Example 3.

TABLE 4A: Nucleotide sequence of POLY3.

GTTTTTTTGTTTTGGAAGAG**ATG**GTCCTGGCTTTCCAGTTAGTCTCCTTCACCTACATCT 60
GGATCATATTGAAACCAAATGTTTGTGCTGCTTCTAACATCAAGATGACACACCAGCGGT 120
GCTCCTCTTCAATGAAACAAACCTGCAAACAAGAAACTAGAATGAAGAAAGATGACAGTA 180
CCAAAGCGCGGCCTCAGAAATATGAGCAACTTCTCCATATAGAGGACAACGATTTCGCAA 240
TGAGACCTGGATTTGGAGGGTCTCCAGTGCCAGTAGGTATAGATGCCCATGTTGAAAGCA 300
TTGACAGCATTTCAGAGACTAACATGGACTTTACAATGACTTTTTATCTCAGGCATTACT 360
GGAAAGACGAGAGGCTCTCCTTTCCTAGCACAGCAAACAAAAGCATGACATTTGATCATA 420
GATTGACCAGAAAGATCTGGGTGCCTGATATCTTTTTTGTCCACTCTAAAAGATCCTTCA 480
TCCATGATACAACTATGGAGAATATCATGCTGCGCGTACACCCTGATGGAAACGTCCTCC 540
TAAGTCTCAGGATAACGGTTTCGGCCATGTGCTTTATGGATTTCAGCAGGTTTCCTCTTG 600
ACGACACTCAAAATTGTTCTCTTGAACTGGAAAGCTGTGCCTACAATGAGGATGACCTAA 660
TGCTATACTGGAAACACGGAAACAAGTCCTTAAATACTGAAGAACATATGTCCCTTTCTC 720
AGTTCTTCATTGAAGACTTCAGTGCATCTAGTGGATTAGCTTTCTATAGCAGCACAGGTT 780
GGTACAATAGGCTTTTCATCAACTTTGTGCTAAGGAGGCATGTTTTCTTCTTTGTGCTGC 840
AAACCTATTTCCCAGCCATATTGATGGTGATGCTTTCATGGGTTTCATTTTGGATTGACC 900
GAAGAGCTGTTCCTGCAAGAGTTTCCCTGGGTATCACCACAGTGCTGACCATGTCCACAA 960
TCATCACTGCTGTGAGCGCCTCCATGCCCCAGGTGTCCTACCTCAAGGCTGTGGATGTGT 1020
ACCTGTGGGTCAGCTCCCTCTTTGTGTTCCTGTCAGTCATTGAGTATGCAGCTGTGAACT 1080
ACCTCACCACAGTGGAAGAGCGGAAACAATTCAAGAAGACAGGAAAGGTATCTAGGATGT 1140
ACAATATTGATGCAGTTCAAGCTATGGCCTTTGATGGTTGTTACCATGACAGCGAGATTG 1200
ACATGGACCAGACTTCCCTCTCTCTAAACTCAGAAGACTTCATGAGAAGAAAATCGATAT 1260
GCAGCCCCAGCACCGATTCATCTCGGATAAAGAGAAGAAAATCCCTAGGAGGACATGTTG 1320
GTAGAATCATTCTGGAAAACAACCATGTCATTGACACCTATTCTAGGATTTTATTCCCCA 1380
TTGTGTATATTTTATTTAATTTGTTTTACTGGGGTGTATATGTAT**GA**AGGGGAATTTCAA 1440
ATGT (SEQ ID NO:5) 1444

TABLE 4B: Protein sequence encoded by the nucleotide sequence shown in TABLE 4A

```
MVLAFQLVSFTYIWIILKPNVCAASNIKMTHQRCSSSMKQTCKQETRMKKDDSTKARPQK    60
YEQLLHIEDNDFAMRPGFGGSPVPVGIDAHVESIDSISETNMDFTMTFYLRHYWKDERLS   120
FPSTANKSMTFDHRLTRKIWVPDIFFVHSKRSFIHDTTMENIMLRVHPDGNVLLSLRITV   180
SAMCFMDFSRFPLDDTQNCSLELESCAYNEDDLMLYWKHGNKSLNTEEHMSLSQFFIEDF   240
SASSGLAFYSSTGWYNRLFINFVLRRHVFFFVLQTYFPAILMVMLSWVSFWIDRRAVPAR   300
VSLGITTVLTMSTIITAVSASMPQVSYLKAVDVYLWVSSLFVFLSVIEYAAVNYLTTVEE   360
RKQFKKTGKVSRMYNIDAVQAMAFDGCYHDSEIDMDQTSLSLNSEDFMRRKSICSPSTDS   420
SRIKRRKSLGGHVGRIILENNHVIDTYSRILFPIVYILFNLFYWGVYV (SEQ ID NO:6)  468
```

In a search of sequence databases, it was found, for example, that the nucleic acid sequence of this invention has 1169 of 1368 bases (85%) identical to a gb:GENBANK-ID:RATGABA|acc:D50671.1 mRNA from Rattus norvegicus (Rat mRNA for GABA receptor rho-3 subunit, complete cds). The full amino acid sequence of the protein of the invention was found to have 390 of 468 amino acid residues (83%) identical to, and 417 of 468 amino acid residues (89%) similar to, the 464 amino acid residue ptnr:SWISSPROT-ACC:P50573 protein from Rattus norvegicus (Rat) (GAMMA-AMINOBUTYRIC-ACID RECEPTOR RHO-3 SUBUNIT PRECURSOR (GABA(A) RECEPTOR))(Table 4C).

TABLE 4C: BLASTX of POLY3 against Gamma-Aminobutyric-Acid Receptor RHO-3 Subunit Precursor (GABA(A)Receptor) (SEQ ID NO:37)

```
>ptnr:SWISSPROT-ACC:P50573 GAMMA-AMINOBUTYRIC-ACID RECEPTOR RHO-3 SUBUNIT
            PRECURSOR (GABA(A) RECEPTOR) - Rattus norvegicus (Rat), 464 aa.
            Length = 464

 Score = 2002 (704.7 bits), Expect = 8.9e-207, P = 8.9e-207
 Identities = 390/468 (83%), Positives = 417/468 (89%)

Query:      1 MVLAFQLVSFTYIWIILKPNVCAASNIKMTHQRCSSSMKQTCKQETRMKKDDSTKARPQK 60
              |||||  |  ||| ||  |   +  || +|  ||+| || |||  +||||+||| ||  | |
Sbjct:      1 MVLAFWLAFFTYTWITL---MLDASAVKEPHQQCLSSPKQTRIRETRMRKDDLTKVWPLK 57

Query:     61 YEQLLHIEDNDFAMRPGFGGSPVPVGIDAHVESIDSISETNMDFTMTFYLRHYWKDERLS 120
              ||||||||+||+ ||||||||||||||  |||||||| ||||||||||||||||||||||
Sbjct:     58 REQLLHIEDHDFSTRPGFGGSPVPVGIDVQVESIDSISEVNMDFTMTFYLRHYWKDERLS 117

Query:    121 FPSTANKSMTFDHRLTRKIWVPDIFFVHSKRSFIHDTTMENIMLRVHPDGNVLLSLRITV 180
              ||||  |||||||  || +|||||||||||||||||||||+||||||||||||||  ||||||
Sbjct:    118 FPSTTNKSMTFDRRLIQKIWVPDIFFVHSKRSFIHDTTVENIMLRVHPDGNVLFSLRITV 177
```

```
Query:   181 SAMCFMDFSRFPLDDTQNCSLELESCAYNEDDLMLYWKHGNKSLNTEEHMSLSQFFIEDF 240
             |||||||||||||| |||||||||| ||||+||||||||||||||||+||||||||+|
Sbjct:   178 SAMCFMDFSRFPLD-TQNCSLELESYAYNEEDLMLYWKHGNKSLNTEEHISLSQFFIEEF 236

Query:   241 SASSGLAFYSSTGWYNRLFINFVLRRHVFFFVLQTYFPAILMVMLSWVSFWIDRRAVPAR 300
             |||||||||||||| ||||||||||+||||||||||+|||||||||||||||||||||||
Sbjct:   237 SASSGLAFYSSTGWYYRLFINFVLRRHIFFFVLQTYFPAMLMVMLSWVSFWIDRRAVPAR 296

Query:   301 VSLGITTVLTMSTIITAVSASMPQVSYLKAVDVYLWVSSLFVFLSVIEYAAVNYLTTVEE 360
             |||||||||||||+| |||||||||+||||||+||||||||||||||||||||||||||
Sbjct:   297 VSLGITTVLTMSTIVTGVSASMPQVSYVKAVDVYMWVSSLFVFLSVIEYAAVNYLTTVEE 356

Query:   361 RKQFKKTGKVSRMYNIDAVQAMAFDGCYHDSEIDMDQTSLSLNSE-DFMRRKSICSPSTD 419
             ||  + ||+| |||||||||||||||||| | |+|||| |+|| | || |   || |
Sbjct:   357 WKQLNRRGKISGMYNIDAVQAMAFDGCYHDGETDVDQTSFFLHSEEDSMRTKFTGSPCAD 416

Query:   420 SSRIKRRKSLGGHVGRIILENNHVIDTYSRILFPIVYILFNLFYWGVYV 468
             ||+||| |||||+|||||||||||||||||+||+|||+|||||||+||
Sbjct:   417 SSQIKR-KSLGGNVGRIILENNHVIDTYSRIVFPVVYIIFNLFYWGIYV 464
```
(SEQ ID NO:37)

Of the five families known, four have been shown to form a sequence-related superfamily. These are the gamma-aminobutyric acid type A (GABA-A), nicotinic acetylcholine, glycine and the serotonin 5HT3 receptors. The ionotropic glutamate receptors have a distinct primary structure. However, all these receptors posess a pentameric structure (made up of varying subunits), surrounding a central pore. Each of these subunits contains a large extracellular N-terminal ligand-binding region; 3 hydrophobic transmembrane domains; a large intracellular region; and a fourth hydrophobic domain. This indicates that the sequence of the invention has properties similar to those of other proteins known to contain this/these domain(s) and similar to the properties of these domains.

PSORT analysis suggests that the GABA Receptor-like protein may be localized in the cytoplasm, although the POLY3 protein (CuraGen Acc. No. CG54683-02) predicted here is similar to the GABA Receptor family, some members of which are membrane localized. Therefore it is likely that this novel GABA Receptor-like protein is localized to the same subcellular compartment.

POLY4

A POLY4 nucleic acid was identified as described in Example 4. A POLY4 nucleic acid was localized to human chromosome 3. The novel nucleic acid of 1438 nucleotides (SEQ

ID NO:7) encoding a novel GABA Receptor-like protein is shown in TABLE 5. An open reading frame was identified beginning at nucleotides 21-23 and ending at nucleotides 1419-1421. The encoded polypeptide represents a novel functional GABA Receptor-like protein (TABLE 5B). The start and stop codons of the open reading frame are highlighted in bold type. Putative untranslated regions (underlined), are found upstream from the initiation codon and downstream from the termination codon. The encoded protein having 466 (SEQ ID NO:8) amino acid residues is presented using the one-letter code in TABLE 5B.

TABLE 5A: Nucleotide sequence of POLY4

GTTTTTTTGTTTTGGAAGAG**ATG**GTCCTGGCTTTCCAGTTAGTCTCCTTCACCTACATCT	60
GGATCATATTGAAACCAAATGTTTGTGCTGCTTCTAACATCAAGATGACACACCAGCGGT	120
GCTCCTCTTCAATGAAACAAACCTGCAAACAAGAAACTAGAATGAAGAAAGATGACAGTA	180
CCAAAGCGCGGCCTCAGAAATATGAGCAACTTCTCCATATAGAGGACAACGATTTCGCAA	240
TGAGACCTGGATTTGGAGGGTCTCCAGTGCCAGTAGGTATAGATGTCCATGTTGAAAGCA	300
TTGACAGCATTTCAGAGACTAACATGGACTTTACAATGACTTTTTATCTCAGGCATTACT	360
GGAAAGACGAGAGGCTCTCCTTTCCTAGCACAGCAAACAAAAGCATGACATTTGATCATA	420
GATTGACCAGAAAGATCTGGGTGCCTGATATCTTTTTTGTCCACTCTAAAAGATCCTTCA	480
TCCATGATACAACTATGGAGAATATCATGCTGCGCGTACACCCTGATGGAAACGTCCTCC	540
TAAGTCTCAGGATAACGGTTTCGGCCATGTGCTTTATGGATTTCAGCAGGTTTCCTCTGA	600
CTCAAAATTGTTCTCTTGAACTGGAAAGCTGTGCCTACAATGAGGATGACCTAATGCTAT	660
ACTGGAAACACGGAAACAAGTCCTTAAATACTGAAGAACATATGTCCCTTTCTCAGTTCT	720
TCATTGAAGACTTCAGTGCATCTAGTGGATTAGCTTTCTATAGCAGCACAGGTTGGTACA	780
ATAGGCTTTTCATCAACTTTGTGCTAAGGAGGCATGTTTTCTTCTTTGTGCTGCAAACCT	840
ATTTCCCAGCCATATTGATGGTGATGCTTTCATGGGTTTCATTTTGGATTGACCGAAGAG	900
CTGTTCCTGCAAGAGTTTCCCTGGGTATCACCACAGTGCTGACCATGTCCACAATCATCA	960
CTGCTGTGAGCGCCTCCATGCCCCAGGTGTCCTACCTCAAGGCTGTGGATGTGTACCTGT	1020
GGGTCAGCTCCCTCTTTGTGTTCCTGTCAGTCATTGAGTATGCAGCTGTGAACTACCTCA	1080
CCACAGTGGAAGAGCGGAAACAATTCAAGAAGACAGGAAAGGTATCTAGGATGTACAATA	1140
TTGATGCAGTTCAAGCTATGGCCTTTGATGGTTGTTACCATGACAGCGAGATTGACATGG	1200
ACCAGACTTCCCTCTCTCTAAACTCAGAAGACTTCATGAGAAGAAAATCGATATGCAGCC	1260
CCAGCACCGATTCATCTCGGATAAAGAGAAGAAAATCCCTAGGAGGACATGTTGGTAGAA	1320
TCATTCTGGAAAACAACCATGTCATTGACACCTATTCTAGGATTTTATTCCCCATTGTGT	1380
ATATTTTATTTAATTTGTTTTACTGGGGTGTATATGTA**TGA**AGGGGAATTTCAAATGT	1438
(SEQ ID NO: 7)	

TABLE 5B:Protein sequence encoded by the nucleotide sequence shown in TABLE 5A

```
MVLAFQLVSFTYIWIILKPNVCAASNIKMTHQRCSSSMKQTCKQETRMKKDDSTKARPQK    60
YEQLLHIEDNDFAMRPGFGGSPVPVGIDVHVESIDSISETNMDFTMTFYLRHYWKDERLS    120
FPSTANKSMTFDHRLTRKIWVPDIFFVHSKRSFIHDTTMENIMLRVHPDGNVLLSLRITV    180
SAMCFMDFSRFPLTQNCSLELESCAYNEDDLMLYWKHGNKSLNTEEHMSLSQFFIEDFSA    240
SSGLAFYSSTGWYNRLFINFVLRRHVFFFVLQTYFPAILMVMLSWVSFWIDRRAVPARVS    300
LGITTVLTMSTIITAVSASMPQVSYLKAVDVYLWVSSLFVFLSVIEYAAVNYLTTVEERK    360
QFKKTGKVSRMYNIDAVQAMAFDGCYHDSEIDMDQTSLSLNSEDFMRRKSICSPSTDSSR    420
IKRRKSLGGHVGRIILENNHVIDTYSRILFPIVYILFNLFYWGVYV  (SEQ ID NO:8)    466
```

In a search of sequence databases, it was found, for example, that the POLY4 nucleic acid sequence of this invention has 1218 of 1430 bases (85%) identical to a gb:GENBANK-ID:RATGABA|acc:D50671.1 mRNA from Rattus norvegicus (Rat mRNA for GABA receptor rho-3 subunit, complete cds). The full amino acid sequence of the protein of the invention was found to have 390 of 466 amino acid residues (83%) identical to, and 417 of 466 amino acid residues (89%) similar to, the 464 amino acid residue ptnr:SWISSPROT-ACC:P50573 protein from Rattus norvegicus (Rat) (GAMMA-AMINOBUTYRIC-ACID RECEPTOR RHO-3 SUBUNIT PRECURSOR (GABA(A) RECEPTOR))(Table 5C).

TABLE 5C: BLASTX of POLY4 against Gamma-Aminobutyric-Acid Receptor RHO-3 Subunit Precursor (GABA(A)Receptor) (SEQ ID NO:38)

```
>ptnr:SWISSPROT-ACC:P50573 GAMMA-AMINOBUTYRIC-ACID RECEPTOR RHO-3 SUBUNIT
           PRECURSOR (GABA(A) RECEPTOR) - Rattus norvegicus (Rat), 464 aa.
           Length = 464

 Score = 2000 (704.0 bits), Expect = 1.5e-206, P = 1.5e-206
 Identities = 390/466 (83%), Positives = 417/466 (89%)

Query:     1 MVLAFQLVSFTYIWIILKPNVCAASNIKMTHQRCSSSMKQTCKQETRMKKDDSTKARPQK 60
             |||||  |  |||  ||  |    +  || +|  ||+|  ||  |||  +||||+|||  ||   | |
Sbjct:     1 MVLAFWLAFFTYTWITL---MLDASAVKEPHQQCLSSPKQTRIRETRMRKDDLTKVWPLK 57

Query:    61 YEQLLHIEDNDFAMRPGFGGSPVPVGIDVHVESIDSISETNMDFTMTFYLRHYWKDERLS 120
             |||||||||+||+ |||||||||||||| |||||||| |||||||||||||||||||||
Sbjct:    58 REQLLHIEDHDFSTRPGFGGSPVPVGIDVQVESIDSISEVNMDFTMTFYLRHYWKDERLS 117

Query:   121 FPSTANKSMTFDHRLTRKIWVPDIFFVHSKRSFIHDTTMENIMLRVHPDGNVLLSLRITV 180
             ||||  |||||||  || +|||||||||||||||||||+|||||||||||||  ||||||
Sbjct:   118 FPSTTNKSMTFDRRLIQKIWVPDIFFVHSKRSFIHDTTVENIMLRVHPDGNVLFSLRITV 177

Query:   181 SAMCFMDFSRFPL-TQNCSLELESCAYNEDDLMLYWKHGNKSLNTEEHMSLSQFFIEDFS 239
             |||||||||||||  |||||||||  ||||+|||||||||||||||||+|||||||+||
```

```
Sbjct:   178 SAMCFMDFSRFPLDTQNCSLELESYAYNEEDLMLYWKHGNKSLNTEEHISLSQFFIEEFS 237

Query:   240 ASSGLAFYSSTGWYNRLFINFVLRRHVFFFVLQTYFPAILMVMLSWVSFWIDRRAVPARV 299
             |||||||||||||| |||||||||||+|||||||||||+||||||||||||||||||||
Sbjct:   238 ASSGLAFYSSTGWYYRLFINFVLRRHIFFFVLQTYFPAMLMVMLSWVSFWIDRRAVPARV 297

Query:   300 SLGITTVLTMSTIITAVSASMPQVSYLKAVDVYLWVSSLFVFLSVIEYAAVNYLTTVEER 359
             |||||||||||||+| ||||||||||+||||||+||||||||||||||||||||||||||
Sbjct:   298 SLGITTVLTMSTIVTGVSASMPQVSYVKAVDVYMWVSSLFVFLSVIEYAAVNYLTTVEEW 357

Query:   360 KQFKKTGKVSRMYNIDAVQAMAFDGCYHDSEIDMDQTSLSLNSE-DFMRRKSICSPSTDS 418
             ||  + ||+| |||||||||||||||||| | |+|||| |+|| | || |   || ||
Sbjct:   358 KQLNRRGKISGMYNIDAVQAMAFDGCYHDGETDVDQTSFFLHSEEDSMRTKFTGSPCADS 417

Query:   419 SRIKRRKSLGGHVGRIILENNHVIDTYSRILFPIVYILFNLFYWGVYV 466
             |+||| |||||+|||||||||||||||||||+||+|||+|||||||+||
Sbjct:   418 SQIKR-KSLGGNVGRIILENNHVIDTYSRIVFPVVYIIFNLFYWGIYV 464
```
(SEQ ID NO:38)

This homology indicates that the sequence of the invention has properties similar to those of other proteins known to contain this/these domain(s) and similar to the properties of these domains.

POLY5-POLY8:

EPIDERMAL GROWTH FACTOR-LIKE NUCLEIC ACIDS AND PROTEINS

POLY5-POLY8 show significant homologies to human (EGF) proteins. It has a characteristic structure with three disulfide bridges, which are essential for its activity. However, many other proteins, including both growth factors and proteins with unrelated functions, have similar EGF-like domains. EGF-like proteins are important in modulation of cell shape, motility, proliferation and differentiation, and are altered in pathologies, *e.g.* cancer, aberrant angiogenesis, renal disease, disorders of the extracellular matrix, and diabetes.

A number of receptor systems have been implicated in playing an important role in the development and progression of many human cancers. The EGF receptor tyrosine kinase family has been found to consistently play a leading role in tumor progression. Indeed, in human breast cancer cases the prognosis of a patient is inversely correlated with the overexpression and/or amplification of this receptor family. Furthermore, downstream

signaling components such as the Src kinases, PI3'K, and the Ras pathway display evidence of deregulation that can accelerate tumor progression. The transgenic mouse system has been ideal in elucidating the biological significance of this receptor family in mammary tumorigenesis. Molecular events involved in mammary tumorigenesis such as ligand binding, receptor dimerization, and the activation of downstream pathways have been addressed using this system. Although there are many molecular steps that appear to drive each stage of tumor development, the EGF receptor family appears to play a causal role in the progression to a transformed phenotype.

POLY5-8 nucleic acids and their encoded polypeptides are useful in a variety of applications and contacts. POLY5-8 are homologous to members of the EGF family of proteins that play an important role in the development and progression of many human cancers.

The expression pattern, and protein similarity information for POLY5-8 suggest that the human EGF -like proteins described herein may function as EGF-like proteins. Therefore, POLY5-8 are useful in potential therapeutic applications implicated in, but not limited to, cancer, and other diseases and disorders. The homology to antigenic secreted and membrane proteins suggests that antibodies directed against the novel genes may be useful in treatment and prevention of cancer, tumorigenesis, and other diseases and disorders.

POLY5-8 are useful in potential therapeutic applications implicated in various diseases and disorders described below and/or other pathologies and disorders. For example, but not limited to, a cDNA encoding the human EGF -like protein may be useful in gene therapy of cancer or other cell proliferative diseases and/or disorders, and the human EGF -like proteins may be useful when administered to a subject in need thereof. By way of non-limiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from, for example, but not limited to, cancer, and other diseases and disorders. POLY5-8 may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed. These materials are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in therapeutic or diagnostic methods.

POLY5

A polynucleotide of the present invention has been identified as clone Z97832_B.0.704 (POLY5). POLY5 is a full-length clone of 4400 nucleotides, including the entire coding sequence of a 735 amino acid protein from nucleotides 679 to 2884. The nucleotide sequence of POLY5 (SEQ ID NO: 9) as presently determined is reported in TABLE 6A. The predicted amino acid sequence of the POLY5 protein (SEQ ID NO: 10) corresponding to the foregoing nucleotide sequence is reported in TABLE 6B.

TABLE 6A: The Nucleotide sequence of POLY5

CACGAGCCGGCTTCCGCCCTCCCCTGGCCGCGAGACCGGCCCCGGCGGCTGGGCCGCCAGTAGCTCCAGCCATGGGCTCG
GGGCGCGTACCCGGGCTCTGCCTGCTTGTCCTGCTGGTCCACGCCCGCGCCGCCCAGTACAGCAAAGCCGCGCAAGATGT
GGATGAGTGTGTGGAGGGGACTGACAACTGCCACATCGATGCTATCTGCCAGAACACCCCGAGGTCATACAAGTGCATCT
GCAAGTCTGGCTACACAGGGGACGGCAAACACTGCAAAGACGTGGATGAGTGCGAGCGAGAGGATAATGCAGGTTGTGTG
CATGACTGTGTCAACATCCCTGGCAATTACCGGTGTACCTGCTATGATGGATTCCACCTGGCACATGACGGACACAACTG
TCTGGATGTGGACGAGTGTGCCGAGGGCAACGGCGGCTGTCAGCAGAGCTGTGTCAACATGATGGGCAGCTATGAGTGCC
ACTGCCGGGAAGGCTTCTTCCTCAGCGACAACCAGCATACCTGTATCCAGCGGCCAGAAGAAGGAATGAATTGCATGAAC
AAGAACCACGGCTGTGCCCACATTTGCCGGGAGACACCCAAGGGGGGTATTGCCTGTGAATGCCGTCCTGGCTTTGAGCT
TACCAAGAACCAACGGGACTGTAAATTGACATGCAACTATGGTAACGGCGGCTGCCAGCACACGTGTGATGACACAGAGC
GAGACCTGTGCTGTCAACAACGGGGGCTGTGACAGTAAGTGCCATGATGCAGCGACTGGTGTCCACTGCACCTGCCCTGT
GGGCTTCATGCTGCAGCCAGACAGGAAGACGTGCAAAGATATAGATGAGTGCCGCTTAAACAACGGGGGCTGTGACCATA
TTTGCCGCAACACAGTGGGCAGCTTCGAATGCAGTTGCAAGAAAGGCTATAAGCTTCTCATCAATGAGAGGAACTGCCAG
GATATAGACGAGTGTTCCTTTGATCGAACCTGTGACCACATATGTGTCAACACACCAGGAAGCTTCCAGTGTCTCTGCCA
TCGTGGCTACCTGTTGTATGGTATCACCCACTGTGGGGATGTGGATGAATGCAGCATCAACCGGGGAGGTTGCCGCTTTG
GCTGCATCAACACTCCTGGCAGCTACCAGTGTACCTGCCCAGCAGGCCAGGGTCGGCTGCACTGGAATGGCAAAGATTGC
ACAGAGCCACTGAAGTGTCAGGGCAGTCCTGGGGCCTCGAAAGCCATGCTCAGCTGCAACCGGTCTGGCAAGAAGGACAC
CTGTGCCCTGACCTGTCCCTCCAGGGCCCGATTTTTGCCAGAGGCTGCAGTGCTGTCCATTAAACAACGGGCCTCCTTCA
AGATCAAGGATGCCAAATGCCGTTTGCACCTGCGAAACAAAGGCAAAACAGAGGAGGCTGGCAGAACCACAGGGCCAGGT
GGTGCCCCCTGCTCTGAATGCCAGGTCACCTTCATCCACCTTAAGTGTGACTCCTCTCGGAAGGGCAAGGGCCGACGGGC
CCGGACCCCTCCAGGCAAAGAGGTCACAAGGCTCACCCTGGAACTGGAGGCAGAGGTCAGAGCCGAAGAAACCACAGCCA
GCTGTGGGCTGCCCTGCCTCCGACAGCGAATGGAACGGCGGCTGAAAGGATCCCTGAAGATGCTCAGAAAGTCCATCAAC
CAGGACCGCTTCCTGCTGCGCCTGGCAGGCCTTGATTATGAGCTGGCCCACAAGCCGGGCCTGGTAGCCGGGGAGCGAGC
AGAGCCGATGGAGTCCTGTAGGCCCGGGCAGCACCGTGCTGGGACCAAGTGTGTCAGCTGCCCGCAGGGAACGTATTACC
ACGGCCAGACGGAGCAGTGTGTGCCATGCCCAGCGGGCACCTTCCAGGAGAGAGAAGGGCAGCTCTCCTGCGACCTTTGC
CCTGGGAGTGATGCCCACGGGCCTCTTGGAGCCACCAACGTCACCACGTGTGCAGGTCAGTGCCCACCTGGCCAACACTC
TGTAGATGGGTTCAAGCCCTGTCAGCCATGCCCACGTGGCACCTACCAACCTGAAGCAGGACGGACCCTATGCTTCCCTT
GTGGTGGGGGCCTCACCACCAAGCATGAAGGGGCCATTTCCTTCCAAGACTGTGACACCAAAGTCCAGTGCTCCCCAGGG
CACTACTACAACACCAGCATCCACCGCTGTATTCGCTGTGCCATGGGCTCCTATCAGCCCGACTTCCGTCAGAACTTCTG
CAGCCGCTGTCCAGGAAACACAAGCACAGACTTTGATGGCTCTACCAGTGTGGCCCAATGCAAGAATCGTCAGTGTGGTG
GGGAGCTGGGTGAGTTCACTGGCTATATTGAGTCCCCCAACTACCCGGGCAACTACCCAGCTGGTGTGGAGTGCATCTGG
AACATCAACCCCCCACCCAAGCGCAAGATCCTTATCGTGGTACCAGAGATCTTCCTGCCATCTGAGGATGAGTGTGGGGA
CGTCCTCGTCATGAGAAAGAACTCATCCCCATCCTCCATTACCACTTATGAGACCTGCCAGACCTACGAGCGTCCCATTG
CCTTCACTGCCCGTTCCAGGAAGCTCTGGATCAACTTCAAGACAAGCGAGGCCAACAGCGCCCGTGGCTTCCAGATTCCC

TATGTTACCTATGATGAGGACTATGAGCAGCTGGTAGAAGACATTGTGCGAGATGGCCGGCTCTATGCCTCTGAAAACCA
CCAGGAGATTTTAAAGGACAAGAAGCTCATCAAGGCCTTCTTTGAGGTGCTAGCCCACCCCCAGAACTACTTCAAGTACA
CAGAGAAACACAAGGAGATGCTGCCAAAATCCTTCATCAAGCTGCTCCGCTCCAAAGTTTCCAGCTTCCTGAGGCCCTAC
AAATAGTAACCCTAGGCTCAGAGACCCAATTTTTTAAGCCCCCAGACTCCTTAGCCCTCAGAGCCGGCAGCCCCCTACCC
TCAGACAAGGAACTCTCTCCTCTCTTTTTGGAGGGAAAAAAAAATATCACTACACAAACCAGGCACTCTCCCTTTCTGTC
TTTCTAGTTTCCTTTCCTTGTCTCTCTCTGCCTGCCTCTCTACTGTTCCCCCTTTTCTAACACACTACCTAGAAAAGCCA
TTCAGTACTGGCTCTAGTCCCCGTGAGATGTAAAGAAACAGTACAGCCCCTTCCACTGCCCATTTTACCAGCTCACATTC
CCGACCCCATCAGCTTGGAAGGGTGCTAGAGGCCCATCAAGGAAGTGGGTCTGGTGGGAAACGGGGAGGGGAAAGAAGGG
CTTCTGCCATTATAGGGTTGTGCCTTGCTAGTCAGGGGCCAAAATGTCCCCTGGCTCTGCTCCCTAGGGTGATTCTAACA
GCCCAGGGTCCTGCCAAAGAAGCCTTTGATTTACAGGCTTAATGCCAGCACCAGTCCTCTGGGGCACATGGTTTGAGCTC
TGGACTTCCCACATGGCCAGCTTTCTTGTCTATACAGATCCTCTCTTTCTTTCCCTACGTCTGCCTGGGGTCTACTCCAT
AAGGGTTTACAAATGGCCCACAACACTGAGTTAGTGGACACCGGCTAAATGAGGAAGAGCAGCAGGCATTGTCATGGTGA
ATGCCCCGCTGTAGCTCCCTGAGAGAAAGACTGTAACTCTGCAGGACAGAAACAAGGTTTTAAAGCATTGCCAAAAAAAA
AGAAAACAGAAAGAAAAAATGTATCATCTAAAGGTCTAGACACAGAACAATTGGAAGTCAACTTCAAACACTAATCCCTT
TTCTTGTCTTCCCTGGCCCAGCCACCTCCTCAGCCCCATGTGATGCTCCCTGGGGGAGCCCTACTCCCCTTGCTACATGT
TGTCCTTAAACATGGTTATTGACCTGAAGCCAGCCTAGGCCTTGCCCTACAGTTGTTTTTCCCTTGTAGCCCCAGCTGGC
TTGTGGGCTTCACCAAAGAGGACCCCACTCTGAAGCCAGCCTGGAGCCACCTACCTCTGGCCTCAGGCTGTGGGCAGCAA
AAGGAATGTGTGTGCACTTGGCGAGCCTCCTGCCCACCCTGTCCACACCTAATAAGTGCAATCATTTTGAGTCTTTCTAT
GTTGTCTAGACGGAGGGGTTTTTGTTTTCTGGGTTTGTTTTTTTGTTTTTGTTTCTTCTTCCTCTATTAGCAAAACCCTAT
TTATAGCTGCCCAAGAGAAAAGAGTGTATGTTTGGAGTGGAAGAAAATCGGTTTTGAATCTCATGAACCTTGAGTGCTGG
AGCATCTGATCTGTCTCTATGCCACCACTGGCCACCTAGAGCCCTTGGCTGTGGTAATCCAGGGTAATTGCGCAGAGGCA
TCTGATGTGTAGGAAAGTAATTCTGGGGATTTGATGGAGCAGAAAGGAGAGAGACCTATGTTTGCTAAACCAATCTTGCT
ATCCCTATGCCTCTCCATGGAGTCAGTGTGGACCTCATGA (SEQ ID NO:9)

TABLE 6B: The Amino Acid sequence of POLY5

MVTAAASTRVMTQSETCAVNNGGCDSKCHDAATGVHCTCPVGFMLQPDRKTCKDIDECRLNNGGCDHICRNTVGSF
ECSCKKGYKLLINERNCQDIDECSFDRTCDHICVNTPGSFQCLCHRGYLLYGITHCGDVDECSINRGGCRFGCINT
PGSYQCTCPAGQGRLHWNGKDCTEPLKCQGSPGASKAMLSCNRSGKKDTCALTCPSRARFLPEAAVLSIKQRASFK
IKDAKCRLHLRNKGKTEEAGRTTGPGGAPCSECQVTFIHLKCDSSRKGKGRRARTPPGKEVTRLTLELEAEVRAEE
TTASCGLPCLRQRMERRLKGSLKMLRKSINQDRFLLRLAGLDYELAHKPGLVAGERAEPMESCRPGQHRAGTKCVS
CPQGTYYHGQTEQCVPCPAGTFQEREGQLSCDLCPGSDAHGPLGATNVTTCAGQCPPGQHSVDGFKPCQPCPRGTY
QPEAGRTLCFPCGGGLTTKHEGAISFQDCDTKVQCSPGHYYNTSIHRCIRCAMGSYQPDFRQNFCSRCPGNTSTDF
DGSTSVAQCKNRQCGGELGEFTGYIESPNYPGNYPAGVECIWNINPPPKRKILIVVPEIFLPSEDECGDVLVMRKN
SSPSSITTYETCQTYERPIAFTARSRKLWINFKTSEANSARGFQIPYVTYDEDYEQLVEDIVRDGRLYASENHQEI
LKDKKLIKAFFEVLAHPQNYFKYTEKHKEMLPKSFIKLLRSKVSSFLRPYK (SEQ ID NO:10)

The predicted amino acid sequences was searched in the publicly available GenBank database. POLY5, as disclosed in this invention, has 72 of 160 amino acids (45%) identical to a *Caenorhabditis elegans* Y64G10A.7 protein (ACC:CAB57911; 1664 aa), and 69 of 169 amino acids (40%) identical and 99 of 169 amino acids (58%) homology to a *Homo sapiens*

fibrillin 5 protein (ACC:CAB56757; 754 aa, fragment). POLY5 also has 273 of 392 amino acids (69%) identical and 322 of 392 amino acids positive to Breast Cancer Protein BCO2 from *Homo sapiens* (Table 6C).

TABLE 6C: BLASTX of POLY5 against Breast Cancer Protein BCO2 (SEQ ID NO:39)

```
Score = 1539 (541.8 bits), Expect = 3.9e-157, P = 3.9e-157
 Identities = 273/392 (69%), Positives = 322/392 (82%), Frame = +1

Query:  1711 LDYELAHKPGLVAGERAEPMESCRPGQHRAGTKCVSCPQGTYYHGQTEQCVPCPAGTFQE 1890
             ++ ++| ||   +  +||   ||  ||  |  +|||| |||| |  |+|+ || ||||
Sbjct:     1 MNLDVAKKPPRTSERQAE---SCGVGQGHAENQCVSCRAGTYYDGARERCILCPNGTFQN 57

Query:  1891 REGQLSCDLCPGSDAHGPLG---ATNVTTCAGQCPPGQHSVDGFKPCQPCPRGTYQPEAG 2061
              |||++|+ ||     | |    | |++ | | | | ++| ||| ||| |   | +|||||
Sbjct:    58 EEGQMTCEPCPRPGNSGALKTPEAWNMSECGGLCQPTEYSADGFAPCQLCALGXFQPEAG 117

Query:  2062 RTLCFPCGGGLTTKHEGAISFQDCDTKVQCSPGHYYNTSIHRCIRCAMGSYQPDFRQNFC 2241
             || |||||||| |||+|| |||||+|+|||||||+|||+ |||||| +|+|||+| +| |
Sbjct:    118 RTSCFPCGGGLATKHQGATSFQDCETRVQCSPGHFYNTTTHRCIRCPVGTYQPEFGKNNC 177

Query:  2242 SRCPGNTSTDFDGSTSVAQCKNRQCGGELGEFTGYIESPNYPGNYPAGVECIWNINPPPK 2421
               |||||+|||||||++ |||||+||||||+|||||||||||||||||  || | ||||||
Sbjct:    178 VSCPGNTTTDFDGSTNITQCKNRRCGGELGDFTGYIESPNYPGNYPANTECTWTINPPPK 237

Query:  2422 RKILIVVPEIFLPSEDECGDVLVMRKNSSPSSITTYETCQTYERPIAFTARSRKLWINFK 2601
             |+|||||||||| ||+||| ||||| || +|+|||||||||||||||+||+|||| ||
Sbjct:    238 RRILIVVPEIFLPIEDDCGDYLVMRKTSSSNSVTTYETCQTYERPIAFTSRSKKLWIQFK 297

Query:  2602 TSEANSARGFQIPYVTYDEDYEQLVEDIVRDGRLYASENHQEILKDKKLIKAFFEVLAHP 2781
             ++| |||||||+||||||||++|+|||||||||||||||||||||||| |+|||||
Sbjct:    298 SNEGNSARGFQVPYVTYDEDYQELIEDIVRDGRLYASENHQEILKDKKLIKALFDVLAHP 357

Query:  2782 QNYFKYT-EKHKEMLPKSFIKLLRSKVSSFLRPYK 2883
             ||||||| ++ +|| |+|||+||||||| ||||||
Sbjct:    358 QNYFKYTAQESREMFPRSFIRLLRSKVSRFLRPYK 392 (SEQ ID NO:39)
```

POLY5 also has high homology to the polypeptides shown in the BLASTX data in Table 6D.

Table 6D. BLASTX alignments of POLY5

Sequences producing High-scoring Segment Pairs:	Reading Frame	High Score	Smallest Sum Prob. P(N)	N

patp:Y07735	Human breast-specific BS200 protein - Homo.+1	1732	1.4e-177	1
patp:B00192	Breast cancer protein BCO2 - Homo sapiens,.+1	1539	3.9e-157	1
patp:B43170	Human ORFX ORF2934 polypeptide sequence SE.+1	682	4.2e-65	1
patp:W99016	Human matrilin-3 - Homo sapiens, 632 aa. +3	382	1.8e-54	2
patp:Y13350	Amino acid sequence of protein PRO219 - Ho.+1	358	2.7e-53	2
patp:Y95340	Human PRO219 antitumour protein - Homo sap.+1	358	2.7e-53	2

PSORT analysis demonstrates that POLY5 is most likely located in the mitochondrial matrix space (certainty=0.7394). SIGNALP analysis is suggests that POLY5 has no N-terminal signal sequence. The predicted molecular weight of POLY5 is 81198.4 daltons.

POLY 6

A polynucleotide of the present invention has been identified as clone Z97832_B.0.707 (POLY6). POLY6 is a full-length clone of 4821 nucleotides, including the entire coding sequence of a 845 amino acid protein from nucleotides 730 to 3265. The nucleotide sequence of POLY6 (SEQ ID NO: 11) as presently determined is reported in TABLE 7A. The predicted amino acid sequence of POLY6 (SEQ ID NO: 12) corresponding to the foregoing nucleotide sequence is reported in TABLE 7B.

TABLE 7A: The Nucleotide sequence of POLY6

TCATGAGGTCCACACTGACTCCATGGAGAGGCATAGGGATAGCAAGATTGGTTTAGCAAACATAGGTCTCTCTCCTTTCT
GCTCCATCAAATCCCCAGAATTACTTTCCTACACATCAGATGCCTCTGCGCAATTACCCTGGATTACCACAGCCAAGGGC
TCTAGGTGGCCAGTGGTGGCATAGAGACAGATCAGATGCTCCAGCACTCAAGGTTCATGAGATTCAAAACCGATTTTCTT
CCACTCCAAACATACACTCTTTTCTCTTGGGCAGCTATAAATAGGGTTTTGCTAATAGAGGAAGAAGAAACAAAAACAAA
AAACAAACCCAGAAAACAAAAACCCCTCCGTCTAGACAACATAGAAAGACTCAAAATGATTGCACTTATTAGGTGTGGAC
AGGGTGGGCAGGAGGCTCGCCAAGTGCACACACATTCCTTTTGCTGCCCACAGCCTGAGGCCAGAGGTAGGTGGCTCCAG
GCTGGCTTCAGAGTGGGGTCCTCTTTGGTGAAGCCCACAAGCCAGCTGGGGCTACAAGGGAAAAACAACTGTAGGGCAAG
GCCTAGGCTGGCTTCAGGTCAATAACCATGTTTAAGGACAACATGTAGCAAGGGGAGTAGGGCTCCCCCAGGGAGCATCA
CATGGGGCTGAGGAGGTGGCTGGGCCAGGGAAGACAAGAAAAGGGATTAGTGTTTGAAGTTGACTTCCAATTGTTCTGTG
TCTAGACCTTTAGATGATACATTTTTTCTTTCTGTTTTCTTTTTTTTTGGCAATGCTTTAAAACCTTGTTTCTGTCCTGC
AGAGTTACAGTCTTTCTCTCAGGGAGCTACAGCGGGGCATTCACCATGACAATGCCTGCTGCTCTTCCTCATTTAGCCGG
TGTCCACTAACTCAGTGTTGTGGGCCATTTGTAAACCCTTATGGAGTAGACCCCAGGCAGACGTAGGGAAAGAAAGAGAG
GATCTGTATAGACAAGAAAGCTGGCCATGTGGGAAGTCCAGAGCTCAAACCATGTGCCCCAGAGGACTGGTGCTGGCATT
AAGCCTGTAAATCAAAGGCTTCTTTGGCAGGACCCTGGGCTGTTAGAATCACCCTAGGGAGCAGAGCCAGGGGACATTTT
GGCCCCTGACTAGCAAGGCACAACCCTATAATGGCAGAAGCCCTTCTTTCCCCTCCCCGTTTCCCACCAGACCCACTTCC
TTGATGGGCCTCTAGCACCCTTCCAAGCTGATGGGGTCGGGAATGTGAGCTGGTAAAATGGGCAGTGGAAGGGGCTGTAC
TGTTTCTTTACATCTCACGGGGACTAGAGCCAGTACTGAATGGCTTTTCTAGGTAGTGTGTTAGAAAAGGGGGAACAGTA
GAGAGGCAGGCAGAGAGAGACAAGGAAAGGAAACTAGAAAGACAGAAAGGGAGAGTGCCTGGTTTGTGTAGTGATATTTT
TTTTTCCCTCCAAAAAGAGAGGAGAGAGTTCCTTGTCTGAGGGTAGGGGGCTGCCGGCTCTGAGGGCTAAGGAGTCTGGG

GGCTTAAAAAATTGGGTCTCTGAGCCTAGGGTTACTATTTGTAGGGCCTCAGGAAGCTGGAAACTTTGGAGCGGAGCAGC
TTGATGAAGGATTTTGGCAGCATCTCCTTGTGTTTCTCTGTGTACTTGAAGTAGTTCTGGGGGTGGGCTAGCACCTCAAA
GAAGGCCTTGATGAGCTTCTTGTCCTTTAAAATCTCCTGGTGGTTTTCAGAGGCATAGAGCCGGCCATCTCGCACAATGT
CTTCTACCAGCTGCTCATAGTCCTCATCATAGGTAACATAGGGAATCTGGAAGCCACGGGCGCTGTTGGCCTCGCTTGTC
TTGAAGTTGATCCAGAGCTTCCTGGAACGGGCAGTGAAGGCAATGGGACGCTCGTAGGTCTGGCAGGTCTCATAAGTGGT
AATGGAGGATGGGGATGAGTTCTTTCTCATGACGAGGACGTCCCCACACTCATCCTCAGATGGCAGGAAGATCTCTGGTA
CCACGATAAGGATCTTGCGCTTGGGTGGGGGGTTGATGTTCCAGATGCACTCCACACCAGCTGGGTAGTTGCCCGGGTAG
TTGGGGGACTCAATATAGCCAGTGAACTCACCCAGCTCCCCACCACACTGACGATTCTTGCATTGGGCCACACTGGTAGA
GCCATCAAAGTCTGTGCTTGTGTTTCCTGGACAGCGGCTGCAGAAGTTCTGACGGAAGTCGGGCTGATAGGAGCCCATGG
CACAGCGAATACAGCGGTGGATGCTGGTGTTGTAGTAGTGCCCTGGGGAGCACTGGACTTTGGTGTCACAGTCTTGGAAG
GAAATGGCCCCTTCATGCTTGGTGGTGAGGCCCCCACCACAAGGGAAGCATAGGGTCCGTCCTGCTTCAGGTTGGTAGGT
GCCACGTGGGCATGGCTGACAGGGCTTGAACCCATCTACAGAGTGTTGGCCAGGTGGGCACTGACCTGCACACGTGGTGA
CGTTGGTGGCTCCAAGAGGCCCGTGGGCATCACTCCCAGGGCAAAGGTCGCAGGAGAGCTGCCCTTCTCTCTCCTGGAAG
GTGCCCGCTGGGCATGGCACACACTGCTCCGTCTGGCCGTGGTAATACGTTCCCTGCGGGCAGCTGACACACTTGGTCCC
AGCACGGTGCTGCCCGGGCCTACAGGACTCCATCGGCTCTGCTCGCTCCCCGGCTACCAGGCCCGGCTTGTGGGCCAGCT
CATAATCAAGGCCTGCCAGGCGCAGCAGGAAGCGGTCCTGGTTGATGGACTTTCTGAGCATCTTCAGGGATCCTTTCAGC
CGCCGTTCCATTCGCTGTCGGAGGCAGGGCAGCCCACAGCTGGCTGTGGTTTCTTCGGCTCTGACCTCTGCCTCCAGTTC
CAGGGTGAGCCTTGTGACCTCTTTGCCTGGAGGGGTCCGGGCCCGTCGGCCCTTGCCCTTCCGAGAGGAGTCACACTTAA
GGTGGATGAAGGTGACCTGGCATTCAGAGCAGGGGGCACCACCTGGCCCTGTGGTTCTGCCAGCCTCCTCTGTTTTGCCT
TTGTTTCGCAGGTGCAAACGGCATTTGGCATCCTTGATCTTGAAGGAGGCCCGTTGTTTAATGGACAGCACTGCAGCCTC
TGGCAAAAATCGGGCCCTGGAGGGACAGGTCAGGGCACAGGTGTCCTTCTTGCCAGACCGGTTGCAGCTGAGCATGGCTT
TCGAGGCCCCAGGACTGCCCTGACACTTCAGTGGCTCTGTGCAATCTTTGCCATTCCAGTGCAGCCGACCCTGGCCTGCT
GGGCAGGTACACTGGTAGCTGCCAGGAGTGTTGATGCAGCCAAAGCGGCAACCTCCCCGGTTGATGCTGCATTCATCCAC
ATCCCCACAGTGGGTGATACCATACAACAGGTAGCCACGATGGCAGAGACACTGGAAGCTTCCTGGTGTGTTGACACATA
TGTGGTCACAGGTTCGATCAAAGGAACACTCGTCTATATCCTGGCAGTTCCTCTCATTGATGAGAAGCTTATAGCCTTTC
TTGCAACTGCATTCGAAGCTGCCCACTGTGTTGCGGCAAATATGGTCACAGCCCCCGTTGTTTAAGCGGCACTCATCTAT
ATCTTTGCACGTCTTCCTGTCTGGCTGCAGCATGAAGCCCACAGGGCAGGTGCAGTGGACACCAGTCGCTGCATCATGGC
ACTTACTGTCACAGCCCCCGTTGTTGACAGCACAGGTCTCATTAGAAACGGCTTGAGTGGGGATGTGCTGCTCTAGCCGC
CTTTCCCCGATGCATGTCTTCCCGTCGGTATGGAGCACAAACTTGATATGGCAGCCGCACCGGGGACCCTGCTCTGTGTC
ATCACACGTGTGCTGGCAGCCGCCGTTACCATAGTTGCATGTCAATTTACAGTCCCGTTGGTTCTTGGTAAGCTCAAAGC
CAGGACGGCATTCACAGGCAATACCCCCCTTGGGTGTCTCCCGGCAAATGTGGGCACAGCCGTGGTTCTTGTTCATGCAA
TTCATTCCTTCTTCTGGCCGCTGGATACAGGTATGCTGGTTGTCGCTGAGGAAGAAGCCTTCCCGGCAGTGGCACTCATA
GCTGCCCATCATGTTGACACAGCTCTGCTGACAGCCGCCGTTGCCCTCGGCACACTCGTCCACATCCAGACAGTTGTGTC
CGTCATGTGCCAGGTGGAATCCATCATAGCAGGTACACCGGTAATTGCCAGGGATGTTGACACAGTCATGCACACAACCT
GCATTATCCTCTCGCTCGCACTCATCCACGTCTTTGCAGTGTTTGCCGTCCCCTGTGTAGCCAGACTTGCAGATGCACTT
GTATGACCTCGGGGTGTTCTGGCAGATAGCATCGATGTGGCAGTTGTCAGTCCCCTCCACACACTCATCCACATCTGGCA
GGGGCAGAGGGGGCACATGAGAACCTCTGTTGGCACCTCTTAAGGGGTGTCTTGAAGGTGGGCTTCCAAGGGCAGAATCC
CCTCTTCTCTAAAACAGAGGCAGTGACCCCCTCCAGAAACAGGTGCTGTCTCACATCTCTCTGATTTCAGAGTAGGCAGA
CACTGATTTTGGGAATTCAGAAGGAACCCCCACTGCCCTCAAAAATACTAAATTCACAGTGACAGCTAAAACTCCATCAT
TCGAAACACTCCTTTTTTTATTTGAAAACAAACAAAAAACCCTTAGAGTGGGTAGTACACTTAACTTGATTAGGAATAAT
CAACTTAAAGTGAATGAGTTTACGGAGAAGGCTTAGAGGGAAAGTTAAGGGAAAAGGCATGGGAACAGTGGTCTCTGGGA
AGGTGGCAGGGTCCAGCAATC (SEQ ID NO:11)

TABLE 7B: The Amino Acid sequence of POLY6

MMGSYECHCREGFFLSDNQHTCIQRPEEGMNCMNKNHGCAHICRETPKGGIACECRPGFELTKNQRDCKLTCNYGN
GGCQHTCDDTEQGPRCGCHIKFVLHTDGKTCIGERRLEQHIPTQAVSNETCAVNNGGCDSKCHDAATGVHCTCPVG

FMLQPDRKTCKDIDECRLNNGGCDHICRNTVGSFECSCKKGYKLLINERNCQDIDECSFDRTCDHICVNTPGSFQC
LCHRGYLLYGITHCGDVDECSINRGGCRFGCINTPGSYQCTCPAGQGRLHWNGKDCTEPLKCQGSPGASKAMLSCN
RSGKKDTCALTCPSRARFLPEAAVLSIKQRASFKIKDAKCRLHLRNKGKTEEAGRTTGPGGAPCSECQVTFIHLKC
DSSRKGKGRRARTPPGKEVTRLTLELEAEVRAEETTASCGLPCLRQRMERRLKGSLKMLRKSINQDRFLLRLAGLD
YELAHKPGLVAGERAEPMESCRPGQHRAGTKCVSCPQGTYYHGQTEQCVPCPAGTFQEREGQLSCDLCPGSDAHGP
LGATNVTTCAGQCPPGQHSVDGFKPCQPCPRGTYQPEAGRTLCFPCGGGLTTKHEGAISFQDCDTKVQCSPGHYYN
TSIHRCIRCAMGSYQPDFRQNFCSRCPGNTSTDFDGSTSVAQCKNRQCGGELGEFTGYIESPNYPGNYPAGVECIW
NINPPPKRKILIVVPEIFLPSEDECGDVLVMRKNSSPSSITTYETCQTYERPIAFTARSRKLWINFKTSEANSARG
FQIPYVTYDEDYEQLVEDIVRDGRLYASENHQEILKDKKLIKAFFEVLAHPQNYFKYTEKHKEMLPKSFIKLLRSK
VSSFLRPYK (SEQ ID NO:12)

POLY6 as disclosed in this invention invention has 112 of 280 amino acids (40%) identical to a *Caenorhabditis elegans* Y64G10A.7 protein (ACC:CAB57911; 1664 aa), and 106 of 304 amino acids (34%) identical and 154 of 304 amino acids (50%) homology to a *Homo sapiens* hypothetical 82.9 kD protein (ACC:CAB70853; 741 aa, fragment).

PSORT analysis demonstrates that POLY6 is most likely located in the nucleus (certainty= 0.7000). SIGNALP analysis suggests that POLY6 has no N-terminal signal sequence. The predicted molecular weight of POLY6 is 93653.3 daltons.

POLY7

A polynucleotide of the present invention has been identified as clone Z97832_B_1 (POLY7). POLY7 is a full-length clone of 4550 nucleotides, including the entire coding sequence of a 974 amino acid protein from nucleotides 72 to 2994. The nucleotide sequence of POLY7 (SEQ ID NO: 13) as presently determined is reported in TABLE 8A. The predicted amino acid sequence of POLY7 (SEQ ID NO: 14) corresponding to the foregoing nucleotide sequence is reported in TABLE 8B.

TABLE 8A: The nucleotide sequence of POLY7

CACGAGCCGGCTTCCGCCCTCCCCTGGCCGCGAGACCGGCCCCGGCGGCTGGGCCGCCAGTAGCTCCAGCCATGGGCTCG
GGGCGCGTACCCGGGCTCTGCCTGCTTGTCCTGCTGGTCCACGCCCGCGCCGCCCAGTACAGCAAAGCCGCGCAAGATGT
GGATGAGTGTGTGGAGGGGACTGACAACTGCCACATCGATGCTATCTGCCAGAACACCCCGAGGTCATACAAGTGCATCT

GCAAGTCTGGCTACACAGGGGACGGCAAACACTGCAAAGACGTGGATGAGTGCGAGCGAGAGGATAATGCAGGTTGTGTG
CATGACTGTGTCAACATCCCTGGCAATTACCGGTGTACCTGCTATGATGGATTCCACCTGGCACATGACGGACACAACTG
TCTGGATGTGGACGAGTGTGCCGAGGGCAACGGCGGCTGTCAGCAGAGCTGTGTCAACATGATGGGCAGCTATGAGTGCC
ACTGCCGGGAAGGCTTCTTCCTCAGCGACAACCAGCATACCTGTATCCAGCGGCCAGAAGAAGGAATGAATTGCATGAAC
AAGAACCACGGCTGTGCCCACATTTGCCGGGAGACACCCAAGGGGGGTATTGCCTGTGAATGCCGTCCTGGCTTTGAGCT
TACCAAGAACCAACGGGACTGTAAATTGACATGCAACTATGGTAACGGCGGCTGCCAGCACACGTGTGATGACACAGAGC
AGGGTCCCCGGTGCGGCTGCCATATCAAGTTTGTGCTCCATACCGACGGGAAGACATGCATCGGGGAAAGGCGGCTAGAG
CAGCACATCCCCACTCAAGCCGTTTCTAATGAGACCTGTGCTGTCAACAACGGGGGCTGTGACAGTAAGTGCCATGATGC
AGCGACTGGTGTCCACTGCACCTGCCCTGTGGGCTTCATGCTGCAGCCAGACAGGAAGACGTGCAAAGATATAGATGAGT
GCCGCTTAAACAACGGGGGCTGTGACCATATTTGCCGCAACACAGTGGGCAGCTTCGAATGCAGTTGCAAGAAAGGCTAT
AAGCTTCTCATCAATGAGAGGAACTGCCAGGATATAGACGAGTGTTCCTTTGATCGAACCTGTGACCACATATGTGTCAA
CACACCAGGAAGCTTCCAGTGTCTCTGCCATCGTGGCTACCTGTTGTATGGTATCACCCACTGTGGGGATGTGGATGAAT
GCAGCATCAACCGGGGAGGTTGCCGCTTTGGCTGCATCAACACTCCTGGCAGCTACCAGTGTACCTGCCCAGCAGGCCAG
GGTCGGCTGCACTGGAATGGCAAAGATTGCACAGAGCCACTGAAGTGTCAGGGCAGTCCTGGGGCCTCGAAAGCCATGCT
CAGCTGCAACCGGTCTGGCAAGAAGGACACCTGTGCCCTGACCTGTCCCTCCAGGGCCCGATTTTTGCCAGAGGCTGCAG
TGCTGTCCATTAAACAACGGGCCTCCTTCAAGATCAAGGATGCCAAATGCCGTTTGCACCTGCGAAACAAAGGCAAAACA
GAGGAGGCTGGCAGAACCACAGGGCCAGGTGGTGCCCCCTGCTCTGAATGCCAGGTCACCTTCATCCACCTTAAGTGTGA
CTCCTCTCGGAAGGGCAAGGGCCGACGGGCCCGGACCCCTCCAGGCAAAGAGGTCACAAGGCTCACCCTGGAACTGGAGG
CAGAGGTCAGAGCCGAAGAAACCACAGCCAGCTGTGGGCTGCCCTGCCTCCGACAGCGAATGGAACGGCGGCTGAAAGGA
TCCCTGAAGATGCTCAGAAAGTCCATCAACCAGGACCGCTTCCTGCTGCGCCTGGCAGGCCTTGATTATGAGCTGGCCCA
CAAGCCGGGCCTGGTAGCCGGGGAGCGAGCAGAGCCGATGGAGTCCTGTAGGCCCGGGCAGCACCGTGCTGGGACCAAGT
GTGTCAGCTGCCCGCAGGGAACGTATTACCACGGCCAGACGGAGCAGTGTGTGCCATGCCCAGCGGGCACCTTCCAGGAG
AGAGAAGGGCAGCTCTCCTGCGACCTTTGCCCTGGGAGTGATGCCCACGGGCCTCTTGGAGCCACCAACGTCACCACGTG
TGCAGGTCAGTGCCCACCTGGCCAACACTCTGTAGATGGGTTCAAGCCCTGTCAGCCATGCCCACGTGGCACCTACCAAC
CTGAAGCAGGACGGACCCTATGCTTCCCTTGTGGTGGGGGCCTCACCACCAAGCATGAAGGGGCCATTTCCTTCCAAGAC
TGTGACACCAAAGTCCAGTGCTCCCCAGGGCACTACTACAACACCAGCATCCACCGCTGTATTCGCTGTGCCATGGGCTC
CTATCAGCCCGACTTCCGTCAGAACTTCTGCAGCCGCTGTCCAGGAAACACAAGCACAGACTTTGATGGCTCTACCAGTG
TGGCCCAATGCAAGAATCGTCAGTGTGGTGGGGAGCTGGGTGAGTTCACTGGCTATATTGAGTCCCCCAACTACCCGGGC
AACTACCCAGCTGGTGTGGAGTGCATCTGGAACATCAACCCCCCACCCAAGCGCAAGATCCTTATCGTGGTACCAGAGAT
CTTCCTGCCATCTGAGGATGAGTGTGGGGACGTCCTCGTCATGAGAAAGAACTCATCCCCATCCTCCATTACCACTTATG
AGACCTGCCAGACCTACGAGCGTCCCATTGCCTTCACTGCCCGTTCCAGGAAGCTCTGGATCAACTTCAAGACAAGCGAG
GCCAACAGCGCCCGTGGCTTCCAGATTCCCTATGTTACCTATGATGAGGACTATGAGCAGCTGGTAGAAGACATTGTGCG
AGATGGCCGGCTCTATGCCTCTGAAAACCACCAGGAGATTTTAAAGGACAAGAAGCTCATCAAGGCCTTCTTTGAGGTGC
TAGCCCACCCCCAGAACTACTTCAAGTACACAGAGAAACACAAGGAGATGCTGCCAAAATCCTTCATCAAGCTGCTCCGC
TCCAAAGTTTCCAGCTTCCTGAGGCCCTACAAATAGTAACCCTAGGCTCAGAGACCCAATTTTTTAAGCCCCCAGACTCC
TTAGCCCTCAGAGCCGGCAGCCCCCTACCCTCAGACAAGGAACTCTCTCCTCTCTTTTTGGAGGGAAAAAAAAATATCAC
TACACAAACCAGGCACTCTCCCTTTCTGTCTTTCTAGTTTCCTTTCCTTGTCTCTCTCTGCCTGCCTCTCTACTGTTCCC
CCTTTTCTAACACACTACCTAGAAAAGCCATTCAGTACTGGCTCTAGTCCCCGTGAGATGTAAAGAAACAGTACAGCCCC
TTCCACTGCCCATTTTACCAGCTCACATTCCCGACCCCATCAGCTTGGAAGGGTGCTAGAGGCCCATCAAGGAAGTGGGT
CTGGTGGGAAACGGGGAGGGGAAAGAAGGGCTTCTGCCATTATAGGGTTGTGCCTTGCTAGTCAGGGGCCAAAATGTCCC
CTGGCTCTGCTCCCTAGGGTGATTCTAACAGCCCAGGGTCCTGCCAAAGAAGCCTTTGATTTACAGGCTTAATGCCAGCA
CCAGTCCTCTGGGGCACATGGTTTGAGCTCTGGACTTCCCACATGGCCAGCTTTCTTGTCTATACAGATCCTCTCTTTCT
TTCCCTACGTCTGCCTGGGGTCTACTCCATAAGGGTTTACAAATGGCCCACAACACTGAGTTAGTGGACACCGGCTAAAT
GAGGAAGAGCAGCAGGCATTGTCATGGTGAATGCCCCGCTGTAGCTCCCTGAGAGAAAGACTGTAACTCTGCAGGACAGA
AACAAGGTTTTAAAGCATTGCCAAAAAAAAAGAAAACAGAAAGAAAAAATGTATCATCTAAAGGTCTAGACACAGAACAA
TTGGAAGTCAACTTCAAACACTAATCCCTTTTCTTGTCTTCCCTGGCCCAGCCACCTCCTCAGCCCCATGTGATGCTCCC
TGGGGGAGCCCTACTCCCCTTGCTACATGTTGTCCTTAAACATGGTTATTGACCTGAAGCCAGCCTAGGCCTTGCCCTAC

AGTTGTTTTTCCCTTGTAGCCCCAGCTGGCTTGTGGGCTTCACCAAAGAGGACCCCACTCTGAAGCCAGCCTGGAGCCAC
CTACCTCTGGCCTCAGGCTGTGGGCAGCAAAAGGAATGTGTGTGCACTTGGCGAGCCTCCTGCCCACCCTGTCCACACCT
AATAAGTGCAATCATTTTGAGTCTTTCTATGTTGTCTAGACGGAGGGGTTTTTGTTTTCTGGGTTTGTTTTTTGTTTTTG
TTTCTTCTTCCTCTATTAGCAAAACCCTATTTATAGCTGCCCAAGAGAAAAGAGTGTATGTTTGGAGTGGAAGAAAATCG
GTTTTGAATCTCATGAACCTTGAGTGCTGGAGCATCTGATCTGTCTCTATGCCACCACTGGCCACCTAGAGCCCTTGGCT
GTGGTAATCCAGGGTAATTGCGCAGAGGCATCTGATGTGTAGGAAAGTAATTCTGGGGATTTGATGGAGCAGAAAGGAGA
GAGACCTATGTTTGCTAAACCAATCTTGCTATCCCTATGCCTCTCCATGGAGTCAGTGTGGACCTCATGA (SE QID NO: 13)

TABLE 8B: The Amino Acid sequence of POLY7

MGSGRVPGLCLLVLLVHARAAQYSKAAQDVDECVEGTDNCHIDAICQNTPRSYKCICKSGYTGDGKHCKDVDECER
EDNAGCVHDCVNIPGNYRCTCYDGFHLAHDGHNCLDVDECAEGNGGCQQSCVNMMGSYECHCREGFFLSDNQHTCI
QRPEEGMNCMNKNHGCAHICRETPKGGIACECRPGFELTKNQRDCKLTCNYGNGGCQHTCDDTEQGPRCGCHIKFV
LHTDGKTCIGERRLEQHIPTQAVSNETCAVNNGGCDSKCHDAATGVHCTCPVGFMLQPDRKTCKDIDECRLNNGGC
DHICRNTVGSFECSCKKGYKLLINERNCQDIDECSFDRTCDHICVNTPGSFQCLCHRGYLLYGITHCGDVDECSIN
RGGCRFGCINTPGSYQCTCPAGQGRLHWNGKDCTEPLKCQGSPGASKAMLSCNRSGKKDTCALTCPSRARFLPEAA
VLSIKQRASFKIKDAKCRLHLRNKGKTEEAGRTTGPGGAPCSECQVTFIHLKCDSSRKGKGRRARTPPGKEVTRLT
LELEAEVRAEETTASCGLPCLRQRMERRLKGSLKMLRKSINQDRFLLRLAGLDYELAHKPGLVAGERAEPMESCRP
GQHRAGTKCVSCPQGTYYHGQTEQCVPCPAGTFQEREGQLSCDLCPGSDAHGPLGATNVTTCAGQCPPGQHSVDGF
KPCQPCPRGTYQPEAGRTLCFPCGGGLTTKHEGAISFQDCDTKVQCSPGHYYNTSIHRCIRCAMGSYQPDFRQNFC
SRCPGNTSTDFDGSTSVAQCKNRQCGGELGEFTGYIESPNYPGNYPAGVECIWNINPPPKRKILIVVPEIFLPSED
ECGDVLVMRKNSSPSSITTYETCQTYERPIAFTARSRKLWINFKTSEANSARGFQIPYVTYDEDYEQLVEDIVRDG
RLYASENHQEILKDKKLIKAFFEVLAHPQNYFKYTEKHKEMLPKSFIKLLRSKVSSFLRPYK (SEQ ID NO:14)

POLY7 as disclosed in this invention has 145 of 355 amino acids (40%) identical to a *Rattus norvegicus* MEGF6 protein (ACC:O88281; 1574 aa), and 151 of 404 amino acids (37%) identical and 205 of 404 amino acids (50%) homology to a *Homo sapiens* hypothetical 82.9 kD protein (ACC:CAB70853; 741 aa, fragment).

PSORT analysis demonstrates that POLY7 is most likely located outside of the cell (certainty= 0.3700). SIGNALP analysis suggests that POLY7 has a cleavable N-term signal sequence with a most likely cleavage site between positions 21 and 22 of SEQ ID NO. 14. The predicted molecular weight of POLY7 is 107538.6 daltons.

POLY 8

A polynucleotide of the present invention has been identified as clone CG55096-04 (POLY8). POLY8 is a full-length clone of 3177 nucleotides, including a coding sequence of a 1009 amino acid protein. The nucleotide sequence of POLY8 (SEQ ID NO: 15) as presently determined is reported in TABLE 9A. The predicted amino acid sequence of POLY8 (SEQ ID NO: 16) corresponding to the foregoing nucleotide sequence is reported in TABLE 9B.

TABLE 9A The nucleotide sequence of POLY8.

```
CCCCTCCCCTCCCCCTCCTGCGAGCTGGGATCCGGCCGGCTTCCGCCCTCCCCTGGCCGCGAGACCGGCC
CCGGCGGCTGGGCCGCCAGTAGCTCCAGCCATGGGCTCGGGGCGCGTACCCGGGCTCTGCCTGCTTGTCC
TGCTGGTCCACGCCCGCGCCGCCCAGTACAGCAAAGCCGCGCAAGATGTGGATGAGTGTGTGGAGGGGAC
TGACAACTGCCACATCGATGCTATCTGCCAGAACACCCCGAGGTCATACAAGTGCATCTGCAAGTCTGGC
TACACAGGGGACGGCAAACACTGCAAAGACGTGGATGAGTGCGAGCGAGAGGATAATGCAGGTTGTGTGC
ATGACTGTGTCAACATCCCTGGCAATTACCGGTGTACCTGCTATGATGGATTCCACCTGGCACATGACGG
ACACAACTGTCTGGATGTGGACGAGTGTGCCGAGGGCAACGGCGGCTGTCAGCAGAGCTGTGTCAACATG
ATGGGCAGCTATGAGTGCCACTGCCGGGAAGGCTTCTTCCTCAGCGACAACCAGCATACCTGTATCCAGC
GGCCAGAAGAAGGAATGAATTGCATGAACAAGAACCACGGCTGTGCCCACATTTGCCGGGAGACACCCAA
GGGGGGTATTGCCTGTGAATGCCGTCCTGGCTTTGAGCTTACCAAGAACCAACGGGACTGTAAATTGACA
TGCAACTATGGTAACGGCGGCTGCCAGCACACGTGTGATGACACAGAGCAGGGTCCCCGGTGCGGCTGCC
ATATCAAGTTTGTGCTCCATACCGACGGGAAGACATGCATCGGGGAAAGGCGGCTAGAGCAGCACATCCC
CACTCAAGCCGTTTCTAATGAGACCTGTGCTGTCAACAACGGGGGCTGTGACAGTAAGTGCCATGATGCA
GCGACTGGTGTCCACTGCACCTGCCCTGTGGGCTTCATGCTGCAGCCAGACAGGAAGACGTGCAAAGATA
TAGATGAGTGCCGCTTAAACAACGGGGGCTGTGACCATATTTGCCGCAACACAGTGGGCAGCTTCGAATG
CAGTTGCAAGAAAGGCTATAAGCTTCTCATCAATGAGAGGAACTGCCAGGATATAGACGAGCGTTCCTTT
GATCGAACCTGTGACCACATATGTGTCAACACACCAGGAAGCTTCCAGTGTCTCTGCCATCGTGGCTACC
TGTTGTATGGTATCACCCACTGTGGGGATGTGGATGAATGCAGCATCAACCGGGGAGGTTGCCGCTTTGG
CTGCATCAACACTCCTGGCAGCTACCAGTGTACCTGCCCAGCAGGCCAGGGTCGGCTGCACTGGAATGGC
AAAGATTGCACAGAGCCACTGAAGTGTCAGGGCAGTCCTGGGGCCTCGAAAGCCATGCTCAGCTGCAACC
GGTCTGGCAAGAAGGACACCTGTGCCCTGACCTGTCCCTCCAGGGCCCGATTTTTGCCAGAGTCTGAGAA
TGGCTTCACGGTGAGCTGTGGGACCCCCAGCCCCAGGGCTGCTCCAGCCCGAGCTGGCCACAATGGGAAC
AGCACCAACTCCAACCACTGCCATGAGGCTGCAGTGCTGTCCATTAAACAACGGGCCTCCTTCAAGATCA
AGGATGCCAAATGCCGTTTGCACCTGCGAAACAAAGGCAAAACAGAGGAGGCTGGCAGAATCACAGGGCC
AGGTGGTGCCCCCTGCTCTGAATGCCAGGTCACCTTCATCCACCTTAAGTGTGACTCCTCTCGGAAGGGC
AAGGGCCGACGGGCCCGGACCCCTCCAGGCAAAGAGGTCACAAGGCTCACCCTGGAACTGGAGGCAGAGG
TCAGAGCCGAAGAAACCACAGCCAGCTGTGGGCTGCCCTGCCTCCGACAGCGAATGGAACGGCGGCTGAA
AGGATCCCTGAAGATGCTCAGAAAGTCCATCAACCAGGACCGCTTCCTGCTGCGCCTGGCAGGCCTTGAT
TATGAGCTGGCCCACAAGCCGGGCCTGGTAGCCGGGGAGCGAGCAGAGCCGATGGAGTCCTGTAGGCCCG
GGCAGCACCGTGCTGGGACCAAGTGTGTCAGCTGCCCGCAGGGAACGTATTACCACGGCCAGACGGAGCA
GTGTGTGCCATGCCCAGCGGGCACCTTCCAGGAGAGAGAAGGGCAGCTCTCCTGCGACCTTTGCCCTGGG
AGTGATGCCCACGGGCCTCTTGGAGCCACCAACGTCACCACGTGTGCAGGTCAGTGCCCACCTGGCCAAC
ACTCTGTAGATGGGTTCAAGCCCTGTCAGCCATGCCCACGTGGCACCTACCAACCTGAAGCAGGACGGAC
CCTATGCTTCCCTTGTGGTGGGGGCCTCACCACCAAGCATGAAGGGGCCATTTCCTTCCAAGACTGTGAC
ACCAAAGTCCAGTGCTCCCCAGGGCACTACTACAACACCAGCATCCACCGCTGTATTCGCTGTGCCATGG
GCTCCTATCAGCCCGACTTCCGTCAGAACTTCTGCAGCCGCTGTCCAGGAAACACAAGCACAGACTTTGA
TGGCTCTACCAGTGTGGCCCAATGCAAGAATCGTCAGTGTGGTGGGGAGCTGGGTGAGTTCACTGGCTAT
ATTGAGTCCCCCAACTACCCGGGCAACTACCCAGCTGGTGTGGAGTGCATCTGGAACATCAACCCCCCAC
CCAAGCGCAAGATCCTTATCGTGGTACCAGAGATCTTCCTGCCATCTGAGGATGAGTGTGGGGACGTCCT
CGTCATGAGAAAGAACTCATCCCCATCCTCCATTACCACTTATGAGACCTGCCAGACCTACGAGCGTCCC
ATTGCCTTCACTGCCCGTTCCAGGAAGCTCTGGATCAACTTCAAGACAAGCGAGGCCAACAGCGCCCGTG
```

GCTTCCAGATTCCCTATGTTACCTATGATGAGGACTATGAGCAGCTGGTAGAAGACATTGTGCGAGATGG
CCGGCTCTATGCCTCTGAAAACCACCAGGAGATTTTAAAGGACAAGAAGCTCATCAAGGCCTTCTTTGAG
GTGCTAGCCCACCCCCAGAACTACTTCAAGTACACAGAGAAACACAAGGAGATGCTGCCAAAATCCTTCA
TCAAGCTGCTCCGCTCCAAAGTTTCCAGCTTCCTGAGGCCCTACAAATAGTAACCCTAGGCTCAGAGACC
CAATTTTTTAAGCCCCCAGACTCCTTA (SEQ ID NO. 15)

TABLE 9B The amino acid sequence of POLY8.

MGSGRVPGLCLLVLLVHARAAQYSKAAQDVDECVEGTDNCHIDAICQNTPRSYKCICKSG
YTGDGKHCKDVDECEREDNAGCVHDCVNIPGNYRCTCYDGFHLAHDGHNCLDVDECAEGN
GGCQQSCVNMMGSYECHCREGFFLSDNQHTCIQRPEEGMNCMNKNHGCAHICRETPKGGI
ACECRPGFELTKNQRDCKLTCNYGNGGCQHTCDDTEQGPRCGCHIKFVLHTDGKTCIGER
RLEQHIPTQAVSNETCAVNNGGCDSKCHDAATGVHCTCPVGFMLQPDRKTCKDIDECRLN
NGGCDHICRNTVGSFECSCKKGYKLLINERNCQDIDERSFDRTCDHICVNTPGSFQCLCH
RGYLLYGITHCGDVDECSINRGGCRFGCINTPGSYQCTCPAGQGRLHWNGKDCTEPLKCQ
GSPGASKAMLSCNRSGKKDTCALTCPSRARFLPESENGFTVSCGTPSPRAAPARAGHNGN
STNSNHCHEAAVLSIKQRASFKIKDAKCRLHLRNKGKTEEAGRITGPGGAPCSECQVTFI
HLKCDSSRKGKGRRARTPPGKEVTRLTLELEAEVRAEETTASCGLPCLRQRMERRLKGSL
KMLRKSINQDRFLLRLAGLDYELAHKPGLVAGERAEPMESCRPGQHRAGTKCVSCPQGTY
YHGQTEQCVPCPAGTFQEREGQLSCDLCPGSDAHGPLGATNVTTCAGQCPPGQHSVDGFK
PCQPCPRGTYQPEAGRTLCFPCGGGLTTKHEGAISFQDCDTKVQCSPGHYYNTSIHRCIR
CAMGSYQPDFRQNFCSRCPGNTSTDFDGSTSVAQCKNRQCGGELGEFTGYIESPNYPGNY
PAGVECIWNINPPPKRKILIVVPEIFLPSEDECGDVLVMRKNSSPSSITTYETCQTYERP
IAFTARSRKLWINFKTSEANSARGFQIPYVTYDEDYEQLVEDIVRDGRLYASENHQEILK
DKKLIKAFFEVLAHPQNYFKYTEKHKEMLPKSFIKLLRSKVSSFLRPYK (SEQ ID NO.
16)

POLY5-8 represent novel members of the EGF family. Based on homology described above, they all share the activity of members of the EGF family and thus are useful in modulating tumor progression.

POLY9-11:

COMPLEMENT RECEPTOR 1 NUCLEIC ACIDS AND PROTEINS

POLY9-11 show significant homology to complement receptor proteins. Complement receptors on neutrophils and eosinophils play a role in activation and adhesion. During asthmatic reactions these receptors have been found elevated on circulating granulocytes. The expression of CD35 (complement receptor type 1) and CD11b (complement receptor type 3) on neutrophils and eosinophils from asthmatic and non-asthmatic children was different. These results indicate that the inducible expression of CD11b on neutrophils and eosinophils

from allergic asthmatic children is primed in vivo.

The expression pattern, and protein similarity information for the invention suggest that the human complement receptor 1-like proteins described in this invention may function as human complement receptor 1-like proteins. Therefore, the nucleic acid and protein of the invention are useful in potential therapeutic applications implicated, for example but not limited to, lung diseases, such as asthma, viral diseases, and other diseases and disorders. The homology to antigenic secreted and membrane proteins suggests that antibodies directed against the novel genes may be useful in treatment and prevention of lung diseases, such as asthma, viral diseases, and other diseases and disorders. and other diseases and disorders.

The nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in lung diseases, such as asthma, viral diseases, and other diseases and disorders. For example, but not limited to, a cDNA encoding the human complement receptor 1-like proteins may be useful in gene therapy for lung diseases such as asthma, and the human complement receptor 1-like proteins may be useful when administered to a subject in need thereof. By way of non-limiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from, for example, but not limited to, asthma, and other diseases and disorders. The novel nucleic acid encoding the human complement receptor 1-like proteins, and the human complement receptor 1-like proteins, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed. These materials are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in therapeutic or diagnostic methods.

POLY9

The novel nucleic acid of 1709 nucleotides, POLY9 (designated CuraGen Acc. No. 10327789.0.16, SEQ ID NO: 17), encodes a novel complement receptor 1-like protein as shown in TABLE 10. A POLY9 nucleic acid is expressed in mammary tissue. An open reading frame was identified beginning with an ATG initiation codon at nucleotide 413 and

ending with a stop codon at nucleotide 1444. The encoded protein having 344 amino acid residues (SEQ ID NO: 18) is presented using the one-letter code in TABLE 10B.

TABLE 10A: The Nucleotide Sequence of POLY9

```
CAGAGTCTTGCTCTGTCTCCCAGGCTGGAGTGCAGTGGCACAATCTCAGCTCACTGCAACCTCTGCCTCCTGGGTTCAAG
TGATTCTCCTGCCTCAGCTTCCCAAATGGCTGAGATTACAGGCACATACCACCATGCCTAGCTAATTTTTGTACAGGTTT
CACCATGTTGGCCAGGCTGGTCTCGAACTCCTAACCTCAAGTGTTCCTCCTGCCTCGGCCTCCCAAAGTGCTGGGATTGT
AGGCATGAATCGTCATGCCGAGCCTAAGTTGACTTTCTACTATCATTTTCACTTATTTAAAAAAATAGAATGGATCTATT
GGAAAAACCATAAATCATTATTTGCTTACTTCCTAATTGATTCATTTTAACATAGACCTTTTAGTTTTTTTCACTATCCA
AGGATTTAGTTAATGCTATCATCTGTTATACAAATCGCACTCACTTGCTTTCTTCCTGTTGCACAGCATACAACTGGCAG
GATCTTTGAGAGTGAAGTGAGGTATCAGTGTAACCCGGGCTATAAGTCAGTCGGAAGTCCTGTATTTGTCTGCCAAGCCA
ATCGCCACTGGCACAGTGAATCCCCTCTGATGTGTGTTCCTCTCGACTGTGGAAAACCTCCCCCGATCCAGAATGGCTTC
ATGAAAGGAGAAAACTTTGAAGTAGGGTCCAAGGTTCAGTTTTTCTGTAATGAGGGTTATGAGCTTGTTGGGGACAGTTC
TTGGACATGTCAGAAATCTGGCAAATGGAATAAGAAGTCAAATCCAAAGTGCATGCCTGCCAAGTGCCCAGAGCCGCCCC
TCTTGGAAAACCAGCTAGTATTAAAGGAGTTGACCACCGAGGTAGGAGTTGTGACATTTTCCTGTAAAGAAGGGCATGTC
CTGCAAGGCCCCTCTGTCCTGAAATGCTTGCCATCCCAGCAATGGAATGACTCTTTCCCTGTTTGTAAGATTGTTCTTTG
TACCCCACCTCCCCTAATTTCCTTTGGTGTCCCCATTCCTTCTTCTGCTCTTCATTTTGGAAGTACTGTCAAGTATTCTT
GTGTAGGTGGGTTTTTCCTAAGAGGAAATTCTACCACCCTCTGCCAACCTGATGGCACCTGGAGCTCTCCACTGCCAGAA
TGTGTTCCAGTAGAATGTCCCCAACCTGAGGAAATCCCCAATGGAATCATTGATGTGCAAGGCCTTGCCTATCTCAGCAC
AGCTCTCTATACCTGCAAGCCAGGCTTTGAATTGGTGGGAAATACTACCACCCTTTGTGGAGAAAATGGTCACTGGCTTG
GAGGAAAACCAACATGTAAAGCCATTGAGTGCCTGAAACCCAAGGAGATTTTGAATGGCAAATTCTCTTACACGGACCTA
CACTATGGACAGACCGTTACCTACTCTTGCAACCGAGGCTTTCGGCTCGAAGGGTCCCAGTGCCTTGACCTGTTTAGAGA
CAGGTGATTGGGATGTAGATGCCCCATCTTGCAATGCCATCCACTGTGATTCCCCACAACCCATTGAAAATGGTTTTGTA
GAAGGTGCAGATTACAGCTATGGTGCCATAATCATCTACAGTTGCTTCCCTGGGTTTCAGGTGGCTGGTCATGCCATGCA
GACCTGTGAAGAGTCAGGATGGTCACTCGTGCCCCCAACATGTATGCCAATAGACTGTGGCCTCCCTCCTCATATAGATT
TTGGAGACTGTACTAAACTCAAAGATGAC  (SEQ ID NO: 17)
```

TABLE 10B: The Amino Acid sequence of POLY9

```
MLSSVIQIALTCFLPVAQHTTGRIFESEVRYQCNPGYKSVGSPVFVCQANRHWHSESPLMCVPLDCGKPPPIQNGF
MKGENFEVGSKVQFFCNEGYELVGDSSWTCQKSGKWNKKSNPKCMPAKCPEPPLLENQLVLKELTTEVGVVTFSCK
EGHVLQGPSVLKCLPSQQWNDSFPVCKIVLCTPPPLISFGVPIPSSALHFGSTVKYSCVGGFFLRGNSTTLCQPDG
TWSSPLPECVPVECPQPEEIPNGIIDVQGLAYLSTALYTCKPGFELVGNTTTLCGENGHWLGGKPTCKAIECLKPK
EILNGKFSYTDLHYGQTVTYSCNRGFRLEGSQCLDLFRDR  (SEQ ID NO:18)
```

In a search of sequence databases, it was found, for example, that POLY9 has 152 of 299 amino acids (35 %) identical to a *Pan troglodytes* (chimpanzee) complement receptor 1 (ACC:Q29530). The full amino acid sequence of the protein of the invention was found to have 105 of 299 amino acid esidues (35 %) identical to, and 152 of 299 residues (50 %) similar to, the 2039 amino acid residue complement receptor 1 from *Homo sapiens* (human) (ACC:Q16745). POLY9 also has homology to a number of other proteins shown in BLASTX data in Table 10C.

Table 10C. BLASTX alignments of POLY9

Sequences producing High-scoring Segment Pairs:	Reading Frame	High Score	Smallest Sum Prob. P(N)	N
atp:B43122 Human ORFX ORF2886 polypeptide sequence SE..+2	+2	481	2.1e-60	2
atp:R36743 CR1 - Homo sapiens, 2039 aa.	+2	414	2.2e-48	3
atp:R11982 Partial human complement type 1 receptor -..+2	+2	412	8.8e-48	3
atp:W45899 Human complememt receptor 1 (residues 1-19..+2	+2	414	1.2e-47	3
atp:Y55751 Human C3b/C4b receptor (CR1) protein - Hom..+2	+2	414	1.5e-47	3
atp:R11810 Human complement type 1 receptor - Homo sa..+2	+2	414	1.5e-47	3

SignalP, Psort and/or hydropathy suggest that POLY9 may be localized outside of the cell (Certainty=0.3700) with a most likely cleavage site between positions 22 and 23 of SEQ ID NO.: 18. Since POLY9 is similar to the "complement receptor family", it is likely that POLY9 is available at the appropriate sub-cellular localization and hence accessible for the therapeutic uses described in this application.

POLY10

A novel nucleic acid of 1952 nucleotides, POLY10 (designated CuraGen Acc. No. 10327789.0.140, SEQ ID NO: 19), encodes a novel complement receptor 1-like protein as shown in TABLE 11A. A POLY10 nucleic acid is expressed in human mammary tissue. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 470 and ending with a stop codon at nucleotide 1687. The encoded protein having 406 amino acid residues (SEQ ID NO: 20) is presented using the one-letter code in TABLE 11B.

TABLE 11A: The Nucleotide sequence of POLY10

TTCAGGAACACCGTCACTTACACTTGCAAAGAAGGCTATACTCTTGCTGGTCTTGACACCATTGAATGCCTGGCCGACGG
CAAGTGGAGTAGAAGTGACCAGCAGTGCCTGGCTGTCTCCTGTGATGAGCCACCCATTGTGGACCACGCCTCTCCAGAGA
CTGCCCATCGGCTCTTTGGAGACATTGCATTCTACTACTGCTCTGATGGTTACAGCCTAGCAGACAATTCCCAGCTTCTC
TGCAATGCCCAGGGCAAGTGGGTACCCCCAGAAGGTCAAGACATGCCCCGTTGTATAGCTCATTTCTGTGAAAAACCTCC
ATCGGTTTCCTATAGCATCTTGGAATCTGTGAGCAAAGCAAAATTTGCAGCTGGCTCAGTTGTGAGCTTTAAATGCATGG
AAGGCTTTGTACTGAACACCTCAGCAAAGATTGAATGTATGAGAGGTGGGCAGTGGAACCCTTCCCCCCATGTCCATCCA
GTGCATCCCTGTGCGGTGTGGAGAGCCACCAAGCATCATGAATGGCTATGCAAGTGGATCAAACTACAGTTTTGGAGCCA
TGGTGGCTTACAGCTGCAACAAGGGGTTCTACATCAAAGGGGAAAAGAAGAGCACCTGCGAAGCCACAGGGCAGTGGAGT
AGTCCTATACCGACGTGCCACCCGGTATCTTGTGGTGAACCACCTAAGGTTGAGAATGGCTTTCTGGAGCATACAACTGG
CAGGATCTTTGAGAGTGAAGTGAGGTATCAGTGTAACCCGGGCTATAAGTCAGTCGGAAGTCCTGTATTTGTCTGCCAAG

CCAATCGCCACTGGCACAGTGAATCCCCTCTGATGTGTGTTCCTCTCGACTGTGGAAAACCTCCCCCGATCCAGAATGGC
TTCATGAAAGGAGAAAACTTTGAAGTAGGGTCCAAGGTTCAGTTTTTCTGTAATGAGGGTTATGAGCTTGTTGGGGACAG
TTCTTGGACATGTCAGAAATCTGGCAAATGGAATAAGAAGTCAAATCCAAAGTGCATGCCTGCCAAGTGCCCAGAGCCGC
CCCTCTTGGAAAACCAGCTAGTATTAAAGGAGTTGACCACCGAGGTAGGAGTTGTGACATTTTCCTGTAAAGAAGGGCAT
GTCCTGCAAGGCCCCTCTGTCCTGAAATGCTTGCCATCCCAGCAATGGAATGACTCTTTCCCTGTTTGTAAGATTGTTCT
TTGTACCCCACCTCCCCTAATTTCCTTTGGTGTCCCCATTCCTTCTTCTGCTCTTCATTTTGGAAGTACTGTCAAGTATT
CTTGTGTAGGTGGGTTTTTCCTAAGAGGAAATTCTACCACCCTCTGCCAACCTGATGGCACCTGGAGCTCTCCACTGCCA
GAATGTGTTCCAGTAGAATGTCCCCAACCTGAGGAAATCCCCAATGGAATCATTGATGTGCAAGGCCTTGCCTATCTCAG
CACAGCTCTCTATACCTGCAAGCCAGGCTTTGAATTGGTGGGAAATACTACCACCCTTTGTGGAGAAAATGGTCACTGGC
TTGGAGGAAAACCAACATGTAAAGCCATTGAGTGCCTGAAACCCAAGGAGATTTTGAATGGCAAATTCTCTTACACGGAC
CTACACTATGGACAGACCGTTACCTACTCTTGCAACCGAGGCTTTCGGCTCGAAGGGTCCCAGTGCCTTGACCTGTTTAG
AGACAGGTGATTGGGATGTAGATGCCCCATCTTGCAATGCCATCCACTGTGATTCCCCACAACCCATTGAAAATGGTTTT
GTAGAAGGTGCAGATTACAGCTATGGTGCCATAATCATCTACAGTTGCTTCCCTGGGTTTCAGGTGGCTGGTCATGCCAT
GCAGACCTGTGAAGAGTCAGGATGGTCACTCGTGCCCCCAACATGTATGCCAATAGACTGTGGCCTCCCTCCTCATATAG
ATTTTGGAGACTGTACTAAACTCAAAGATGAC (SEQ ID NO:19)

TABLE 11B: The Amino Acid sequence of POLY10

MSIQCIPVRCGEPPSIMNGYASGSNYSFGAMVAYSCNKGFYIKGEKKSTCEATGQWSSPIPTCHPVSCGEPPKVEN
GFLEHTTGRIFESEVRYQCNPGYKSVGSPVFVCQANRHWHSESPLMCVPLDCGKPPPIQNGFMKGENFEVGSKVQF
FCNEGYELVGDSSWTCQKSGKWNKKSNPKCMPAKCPEPPLLENQLVLKELTTEVGVVTFSCKEGHVLQGPSVLKCL
PSQQWNDSFPVCKIVLCTPPPLISFGVPIPSSALHFGSTVKYSCVGGFFLRGNSTTLCQPDGTWSSPLPECVPVEC
PQPEEIPNGIIDVQGLAYLSTALYTCKPGFELVGNTTTLCGENGHWLGGKPTCKAIECLKPKEILNGKFSYTDLHY
GQTVTYSCNRGFRLEGSQCLDLFRDR (SEQ ID NO:20)

In a search of sequence databases, it was found, for example, that POLY10 has 101 of 298 amino acids (33 %) identical to a *Pan troglodytes* (chimpanzee) complement receptor 1 (ACC:Q29530). The full amino acid sequence of POLY10 was found to have 105 of 299 amino acid residues (35 %) identical to, and 152 of 299 residues (50 %) similar to, the 2039 amino acid residue complement receptor type 1 precursor (C3B/C4B receptor) (CD35 antigen) from *Homo sapiens* (human) (ACC:P17927). SignalP, Psort and/or hydropathy suggest that the protein may be localized in the cytoplasm of the cell (Certainty=0.6500). Since POLY10 is similar to the "complement receptor family", it is likely that POLY10 is available at the appropriate sub-cellular localization and hence accessible for the therapeutic uses described in this application.

POLY11

A novel nucleic acid of 6153 nucleotides, POLY11 (designated CuraGen Acc. No. 10327789_1, SEQ ID NO: 21) encodes a novel human complement receptor-like protein as shown in TABLE 12A. APOLY11 nucleic acid is expressed in the following tissues: mammary gland, hypothalamus, lymph node, fetal liver, pooled adrenal gland/placenta, placenta, cervix, testicular tumor, adipose, ovary, ascending colon, lymph node, bone marrow, stomach, and fetal lung. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 1-3 and ending with a stop codon at nucleotides 6151-6153. The encoded protein having 2050 amino acid residues (SEQ ID NO: 22) is presented using the one-letter code in TABLE 12B. The predicted molecular weight of the protein is 224498.2 Da.

TABLE 12A: The Nucleotide sequence of POLY11

ATGGCGGGCGCCCCTCCCCCAGCCTCGTTGCCGCCTTGCAGTTTGATCTCAGACTGCTGTGCTAGCAATCAGCGAG
ATTCCGTGGGCGTAGGACCCTCTGAGCCAGGTGTGGGATATAGTCTCGTGGTGCGCCGTTTCTTAAGCCGGTCTGA
AAAGCGCAATATTCGGGTGGGAGTGACCCGATTTTCCAGCTATACTCTTGCTGGTCTTGACACCATTGAATGCCTG
GCCGACGGCAAGTGGAGTAGAAGTGACCAGCAGTGCCTGGCTGTCTCCTGTGATGAGCCACCCATTGTGGACCACG
CCTCTCCAGAGACTGCCCATCGGCTCTTTGGAGACATTGCATTCTACTACTGCTCTGATGGTTACAGCCTAGCAGA
CAATTCCCAGCTTCTCTGCAATGCCCAGGGCAAGTGGGTACCCCCAGAAGGTCAAGACATGCCCCGTTGTATAGCT
CATTTCTGTGAAAAACCTCCATCGGTTTCCTATAGCATCTTGGAATCTGTGAGCAAAGCAAAATTTGCAGCTGGCT
CAGTTGTGAGCTTTAAATGCATGGAAGGCTTTGTACTGAACACCTCAGCAAAGATTGAATGTATGAGAGGTGGGCA
GTGGAACCCTTCCCCCATGTCCATCCAGTGCATCCCTGTGCGGTGTGGAGAGCCACCAAGCATCATGAATGGCTAT
GCAAGTGGATCAAACTACAGTTTTGGAGCCATGGTGGCTTACAGCTGCAACAAGGGGTTCTACATCAAAGGGGAAA
AGAAGAGCACCTGCGAAGCCACAGGGCAGTGGAGTAGTCCTATACCGACGTGCCACCCGGTATCTTGTGGTGAACC
ACCTAAGGTTGAGAATGGCTTTCTGGAGCATACAACTGGCAGGATCTTTGAGAGTGAAGTGAGGTATCAGTGTAAC
CCGGGCTATAAGTCAGTCGGAAGTCCTGTATTTGTCTGCCAAGCCAATCGCCACTGGCACAGTGAATCCCCTCTGA
TGTGTGTTCCTCTCGACTGTGGAAAACCTCCCCCGATCCAGAATGGCTTCATGAAAGGAGAAAACTTTGAAGTAGG
GTCCAAGGGTCAGTTTTTCTGTAATGAAGGGTTATNGAGCTTTGTTGGGGACAGTTCTTGGACATGTCAGAAATCT
GGCAAATGGAATAAGAAGTCAAATCCAAAGTGCATGCCTGCCAAGTGCCCAGAGCCGCCCCTCTTGGAAAACCAGC
TAGTATTAAAGGAGTTGACCACCGAGGTAGGAGTTGTGACATTTTCCTGTAAAGAAAGGCATGTCCTGCAAGGCCC
CTCTGTCCTGAAATGCTTGCCATCCCAGCAATGGAATGACTCTTTCCCTGTTTGTAAGATTGTTCTTTGTACCCCA
CCTCCCCTAATTTCCTTTGGTGTCCCCATTCCTTCTTCTGCTCTTCATTTTGGAAGTACTGTCAAGTATTCTTGTG
TAGGTGGGTTTTTCCTAAGAGGAAATTCTACCACCCTCTGCCAACCTGATGGCACCTGGAGCTCTCCACTGCCAGA
ATGTGTTCCAGTAGAATGTCCCCAACCTGAGGAAATCCCCAATGGAATCATTGATGTGCAAGGCCTTGCCTATCTC
AGCACAGCTCTCTATACCTGCAAGCCAGGCTTTGAATTGGTGGGAAATACTACCACCCTTTGTGGAGAAAATGGTC
ACTGGCTTGGAGGAAAACCAACATGTAAAGCCATTGAGTGCCTGAAACCCAAGGAGATTTTGAATGGCAAATTCTC
TTACACGGACCTACACTATGGACAGACCGTTACCTACTCTTGCAACCGAGGCTTTCGGCTCGAAGGTCCCAGTGCC

TTGACCTGTTTAGAGACAGGTGATTGGGATGTAGATGCCCCATCTTGCAATGCCATCCACTGTGATTCCCCACAAC
CCATTGAAAATGGTTTTGTAGAAGGTGCAGATTACAGCTATGGTGCCATAATCATCTACAGTTGCTTCCCTGGGTT
TCAGGTGGCTGGTCATGCCATGCAGACCTGTGAAGAGTCAGGATGGTCAAGTTCCATCCCAACATGTATGCCAATA
GACTGTGGCCTCCCTCCTCATATAGATTTTGGAGACTGTACTAAACTCAAAGATGACCAGGGATATTTTGAGCAAG
AAGACGACATGATGGAAGTTCCATATGTGACTCCTCACCCTCCTTATCATTTGGGAGCAGTGGCTAAAACCTGGGA
AAATACAAAGGAGTCTCCTGCTACACATTCATCAAACTTTCTGTATGGTACCATGGTTTCATACACCTGTAATCCA
GGATATGAACTTCTGGGGAACCCTGTGCTGATCTGCCAGGAAGATGGAACTTGGAATGGCAGTGCACCATCCTGCA
TTTCAATTGAATGTGACTTGCCTACTGCTCCTGAAAATGGCTTTTTGCGTTTTACAGAGACTAGCATGGGAAGTGC
TGTGCAGTATAGCTGTAAACCTGGACACATTCTAGCAGGCTCTGACTTAAGGCTTTGTCTAGAGAATAGAAAGTGG
AGTGGTGCCTCCCCACGCTGTGAAGCCATTTCATGCAAAAAGCCAAATCCAGTCATGAATGGATCCATCAAAGGAA
GCAACTACACATACCTGAGCACGTTGTACTATGAGTGTGACCCCGGATATGTGCTGAATGGCACTGAGAGGAGAAC
ATGCCAGGATGACAAAAACTGGGATGAGGATGAGCCCATTTGCATTCCTGTGGACTGCAGTTCACCCCCAGTCTCA
GCCAATGGCCAGGTGAGAGGAGACGAGTACACATTCCAAAAAGAGATTGAATACACTTGCAATGAAGGGTTCTTGC
TTGAGGGAGCCAGGAGTCGGGTTTGTCTTGCCAATGGAAGTTGGAGTGGAGCCACTCCCGACTGTGTGCCTGTCAG
ATGTGCCACCCCGCCACAACTGGCCAATGGGGTGACGGAAGGCCTGGACTATGGCTTCATGAAGGAAGTAACATTC
CACTGTCACGAGGGCTACATCTTGCACGGTGCTCCAAAACTCACCTGTCAGTCAGATGGCAACTGGGATGCAGAGA
TTCCTCTCTGTAAACCAGTCAACTGTGGACCTCCTGAAGATCTTGCCCATGGTTTCCCTAATGGTTTTTCCTTTAT
TCATGGGGGCCATATACAGTATCAGTGCTTTCCTGGTTATAAGCTCCATGGAAATTCATCAAGAAGGTGCCTCTCC
AATGGCTCCTGGAGTGGCAGCTCACCTTCCTGCCTGCCTTGCAGATGTTCCACACCAGTAATTGAATATGGAACTG
TCAATGGGACAGATTTTGACTGTGGAAAGGCAGCCCGGATTCAGTGCTTCAAAGGCTTCAAGCTCCTAGGACTTTC
TGAAATCACCTGTGAAGCCGATGGCCAGTGGAGCTCTGGGTTCCCCCACTGTGAACACACTTCTTGTGGTTCTCTT
CCAATGATACCAAATGCGTTCATCAGTGAGACCAGCTCTTGGAAGGAAAATGTGATAACTTACAGCTGCAGGTCTG
GATATGTCATACAAGGCAGTTCAGATCTGATTTGTACAGAGAAAGGGGTATGGAGCCAGCCTTATCCAGTCTGTGA
GCCCTTGTCCTGTGGGTCCCCACCGTCTGTCGCCAATGCAGTGGCAACTGGAGAGGCACACACCTATGAAAGTGAA
GTGAAACTCAGATGTCTGGAAGGTTATACGATGGATACAGATACAGATACATTCACCTGTCAGAAAGATGGTCGCT
GGTTCCCTGAGAGAATCTCCTGCAGTCCTAAAAAATGTCCTCTCCCGGAAAACATAACACATATACTTGTACATGG
GGACGATTTCAGTGTGAATAGGCAAGTTTCTGTGTCATGTGCAGAAGGGTATACCTTTGAGGGAGTTAACATATCA
GTATGTCAGCTTGATGGAACCTGGGAGCCACCATTCTCCGATGAATCTTGCAGTCCAGTTTCTTGTGGGAAACCTG
AAAGTCCAGAACATGGATTTGTGGTTGGCAGTAAATACACCTTTGAAAGCACAATTATTTATCAGTGTGAGCCTGG
CTATGAACTAGAGAATTTGGCTGTGAATCCATCTGGTCCTGGACTTTTCTTGGTTGACAGGACCCTCAGCTGCAGG
TCGGAGTTGGCTAGAGGTCCAATCCAGACCCTGTTTGCCTGGGTATCAGCAGCAGAGGGTGCAGAACAGCGGATAT
TGGTGAACCGCAAATGCTGCTGCCTGATCATTCCTCTGGAAGTTTTGTCTCAGAGGAATACCCGGCCATGTGAGGT
GTCAGTCCGCCCCTACTGGGGGGGGAACAGGGAACGTGTCTGCCAGGAGAACAGACAGTGGAGTGGAGGGGTGGCA
ATATGCAAAGAGACCAGGTGTGAAACTCCACTTGAATTTCTCAATGGGAAAGCTGACATTGAAAACAGGACGACTG
GACCCAACGTGGTATATTCCTGCAACAGAGGCTACAGTCTTGAAGGGCCATCTGAGGCACACTGCACAGAAAATGG
AACCTGGAGCCACCCAGTCCCTCTCTGCAAACCAAATCCATGCCCTGTTCCTTTTGTGATTCCCGAGAATGCTCTG
CTGTCTGAAAAGGAGTTTTATGTTGATCAGAATGTGTCCATCAAATGTAGGGAAGGTTTTCTGCTGCAGGGCCACG
GCATCATTACCTGCAACCCCGACGAGACGTGGACACAGACAAGCGCCAAATGTGAAAGAAGATATACACAACAGCC

CAAGTCCCTGAATTTTCAGCTAGCAGCTTATTGCAGTATTAGAATGTTTATTTTGCGGGGAGGGGTTCAAGATGGC
CAACTAGAAACAGCTGTGGCCGGAGCCTCCCACCGAGAAGAACAAAAACAAAAGCGAGAAAAAGCAAGGTGGTACA
ACGGCCCACCTGGGAGCCACATGGGGCAAGCAGAGCTCCCACCCCCAGCCAAAGGAGGTGGACCTCCCTGCGGGAA
TTTCAGCAACTCCAGCCAGGGGTTTATGAACAGACCTCTGATCTCCCTGAGATGGAGCCCCTGGGGCTCCATGTGG
CCATGGTCTCCACAGATCAGCAGGCTTAGTCCTTCCCCTGCTGGCTCTGAGGAATCCAGGCAGGCTGGACTAGTGG
GATTCCCCACAGCACAGTTTACCTGCTCTGCCAAGGGGCAGCTAGAGCGCTTTGTTAAGCGAGTCCCTGATCCCAT
GCCTCCTGATTGGGATGAGACCCCCCCACAACAGGGGTCACGGATGAGACCCCCCCACAACAGGGGTCACCAGACA
CCTTATACAAGGGTGTTCCTGCTAGCATCAGGTCAGTGCCCCTCTGGGACAGAGCTCCCAGAGGAAAGAGCAGGCA
GCCATCTTTGCTGTTCTGCAGCGTCCGCTGGAAAAGCACAGAATTGGGCAGAGGCTAGGATTGATGAATTGAAAGA
AGTAGGCTTCAGAAAGTGGGTAATAATGAAGTTCGCTGAGCTAAAGGAACATGTTCTAAACCAATGCAAAGACGCC
AAGAACCAGGATAAAACATTACAGGATCCGTTAACCAGAATAACCAGTTTAGAAAGGAATGTAAATGACCTGATGG
AGCTGAAAAACACAACACGAGAACTTCACAATGCAACAACAAAACAAGGCCAACATTCCAGTTCAGGAAATCCAGA
GAACCCCAGTAAGATACTCCATGAGAAGATCAACCCCAAGACACATAATCCTCAGGTTCTCCAAGAAATCTCATGT
GGTCCACCAGCTCACGTAGAAAATGCAATTGCTCGAGGCGTACATTATCAATATGGAGACATGATCACCTACTCAT
GTTACAGTGGATACATGTTGGAGGGTTTCCTGAGGAGTGTTTGTTTAGAAAATGGAACATGGACATCACCTCCTAT
TTGCAGAGCTGTCTGTCGATTTCCATGTCAGAATGGGGGCATCTGCCAACGCCCAAATGCTTGTTCCTGTCCAGAG
GGCTGGATGGGGCGCCTCTGTGAAGAACCAATCTGCATTCTTCCCTGTCTGAACGGAGGTCGCTGTGTGGCCCCTT
ACCAGTGTGACTGCCCGCCTGGCTGGACGGGGTCTCGCTGTCATACAGGTAGGCCTCTTTCATGGTTTGTTTTCTT
GGTGGCTCAGGCCCATGAAACTCCAGAGGACATTGAAGAGTGTGACTTAGACTCAGAAGTGGTGGCAAAATGA
(SEQ ID NO:21)

TABLE 12B: The Amino Acid sequence of POLY11

MAGAPPPASLPPCSLISDCCASNQRDSVGVGPSEPGVGYSLVVRRFLSRSEKRNIRVGVTRFSSYTLAGLDTIECL
ADGKWSRSDQQCLAVSCDEPPIVDHASPETAHRLFGDIAFYYCSDGYSLADNSQLLCNAQGKWVPPEGQDMPRCIA
HFCEKPPSVSYSILESVSKAKFAAGSVVSFKCMEGFVLNTSAKIECMRGGQWNPSPMSIQCIPVRCGEPPSIMNGY
ASGSNYSFGAMVAYSCNKGFYIKGEKKSTCEATGQWSSPIPTCHPVSCGEPPKVENGFLEHTTGRIFESEVRYQCN
PGYKSVGSPVFVCQANRHWHSESPLMCVPLDCGKPPPIQNGFMKGENFEVGSKGQFFCNEGLXSFVGDSSWTCQKS
GKWNKKSNPKCMPAKCPEPPLLENQLVLKELTTEVGVVTFSCKERHVLQGPSVLKCLPSQQWNDSFPVCKIVLCTP
PPLISFGVPIPSSALHFGSTVKYSCVGGFFLRGNSTTLCQPDGTWSSPLPECVPVECPQPEEIPNGIIDVQGLAYL
STALYTCKPGFELVGNTTTLCGENGHWLGGKPTCKAIECLKPKEILNGKFSYTDLHYGQTVTYSCNRGFRLEGPSA
LTCLETGDWDVDAPSCNAIHCDSPQPIENGFVEGADYSYGAIIIYSCFPGFQVAGHAMQTCEESGWSSSIPTCMPI
DCGLPPHIDFGDCTKLKDDQGYFEQEDDMMEVPYVTPHPPYHLGAVAKTWENTKESPATHSSNFLYGTMVSYTCNP
GYELLGNPVLICQEDGTWNGSAPSCISIECDLPTAPENGFLRFTETSMGSAVQYSCKPGHILAGSDLRLCLENRKW
SGASPRCEAISCKKPNPVMNGSIKGSNYTYLSTLYYECDPGYVLNGTERRTCQDDKNWDEDEPICIPVDCSSPPVS
ANGQVRGDEYTFQKEIEYTCNEGFLLEGARSRVCLANGSWSGATPDCVPVRCATPPQLANGVTEGLDYGFMKEVTF
HCHEGYILHGAPKLTCQSDGNWDAEIPLCKPVNCGPPEDLAHGFPNGFSFIHGGHIQYQCFPGYKLHGNSSRRCLS
NGSWSGSSPSCLPCRCSTPVIEYGTVNGTDFDCGKAARIQCFKGFKLLGLSEITCEADGQWSSGFPHCEHTSCGSL

PMIPNAFISETSSWKENVITYSCRSGYVIQGSSDLICTEKGVWSQPYPVCEPLSCGSPPSVANAVATGEAHTYESE
VKLRCLEGYTMDTDTDTFTCQKDGRWFPERISCSPKKCPLPENITHILVHGDDFSVNRQVSVSCAEGYTFEGVNIS
VCQLDGTWEPPFSDESCSPVSCGKPESPEHGFVVGSKYTFESTIIYQCEPGYELENLAVNPSGPGLFLVDRTLSCR
SELARGPIQTLFAWVSAAEGAEQRILVNRKCCCLIIPLEVLSQRNTRPCEVSVRPYWGGNRERVCQENRQWSGGVA
ICKETRCETPLEFLNGKADIENRTTGPNVVYSCNRGYSLEGPSEAHCTENGTWSHPVPLCKPNPCPVPFVIPENAL
LSEKEFYVDQNVSIKCREGFLLQGHGIITCNPDETWTQTSAKCERRYTQQPKSLNFQLAAYCSIRMFILRGGVQDG
QLETAVAGASHREEQKQKREKARWYNGPPGSHMGQAELPPPAKGGGPPCGNFSNSSQGFMNRPLISLRWSPWGSMW
PWSPQISRLSPSPAGSEESRQAGLVGFPTAQFTCSAKGQLERFVKRVPDPMPPDWDETPPQQGSRMRPPHNRGHQT
PYTRVFLLASGQCPSGTELPEERAGSHLCCSAASAGKAQNWAEARIDELKEVGFRKWVIMKFAELKEHVLNQCKDA
KNQDKTLQDPLTRITSLERNVNDLMELKNTTRELHNATTKQGQHSSSGNPENPSKILHEKINPKTHNPQVLQEISC
GPPAHVENAIARGVHYQYGDMITYSCYSGYMLEGFLRSVCLENGTWTSPPICRAVCRFPCQNGGICQRPNACSCPE
GWMGRLCEEPICILPCLNGGRCVAPYQCDCPPGWTGSRCHTGRPLSWFVFLVAQAHETPEDIEECDLDSEVVAK
(SEQ ID NO:22)

In a search of sequence databases, it was found, for example, that the amino acid sequence of POLY11 has 136 of 427 amino acid residues (31%) identical and 195 of 427 amino acid residues (45%) positive with the 2489 amino acid residue human complement receptor-1 (ACC:Q16744).

The POLY11 sequence disclosed in the present application includes within it shorter sequences having nearly 100% identity of the nucleotide sequence in the respective comparisons (see POLY10 and POLY9) and SeqListing#6077 in US Ser. No. 09/540,763 filed March 30, 2000. SignalP, Psort and/or hydropathy suggest that the protein may be localized in the cytoplasm with a certainty=0.4500. No signal peptide is predicted. The expression pattern of a POLY11 nucleic acid in various cells and tissues described in Example 5; POLY11 is highly expressed in normal adipose, cerebellum, lung, mammary gland and placenta.

POLY12: HEMATOPOIETIC STEM AND PROGENITOR CELLS (HSPC)

POLY12 shows significant homology to hematopoietic stem and proginitor cells (HSPC).

The direct effects of cyclosporin A on proliferation of hematopoietic stem and progenitor cells (HSPC)have been well known. Perry et al., Cell Transplant 1999 Jul-Aug;8(4):339-44. Cyclosporin A (Cy A) has been reported to both stimulate and inhibit bone marrow colony assays in a dose-dependent manner. The observation that anti-gamma-IFN antibodies stimulate hematopoiesis to the same degree as Cy A has led several groups to propose that the stimulatory effects of Cy A are due to inhibition of gamma-IFN production by T cells. They show that Cy A can stimulate hematopoietic stem cell growth independent of mediation by T cells. Consequently, these results argue for a direct positive effect of Cy A on the signal transduction pathways in HSPC. There is also direct evidence of selectin ligands on HSPC under physiologic flow conditions and are the first to show a correlation between the maturity of HSPC during development and rolling efficiency on selectins, suggesting a mechanism by which HSPC subsets may differentially home to the extravascular spaces of the bone marrow. Furthermore, it has been discovered that cell cycle progression by itself cannot account for the decrease in repopulating potential that is observed after ex vivo expansion. Other determinants of engraftment must be identified to facilitate the transplantation of cultured HSPC.

The POLY12 nucleic acid and its encoded polypeptide are useful in a variety of applications and contexts. POLY12 is homologous to members of the HSPC family of proteins that are important in hematopoietic cell proliferation and differentiation. Therefore, POLY12 nucleic acids, proteins, antibodies and other compositions of the present invention are useful in diagnostic and therapeutic applications in disorders of the hematopoietic system, e.g. leukemia, systemic lupus erythematosus, and chronic aplastic anemia. POLY12 also has utility as a marker for alterations in gene expression in cells following cyclosporin A and gamma-interferon.d

POLY12

The novel nucleic acid of 2216 nucleotides, POLY12 (designated CuraGen Acc. No. AC016030_A.0.82) encodes a novel hematopoietic stem and progenitor cell-like protein as shown in TABLE 13A. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 720 and ending with a stop codon at nucleotides 2090. The

encoded protein having 457 amino acid residues is presented using the one-letter code in TABLE 13B.

TABLE 13A: The Nucleotide sequence of POLY12

CCCACGCGTCCGCCCACGCGTCCGCCCACGCGTCCGCCCACGCGTCCGCCCACGCGTCCG
CCCACGCGTCCGCCCACGCGTCCGGTGCAAGCTCGCGCCGCACACTGCCTGGTGGAGGGA
AGGAGCCCGGGCGCCTCTCGCCGCTCCCCGCGCCGCCGTCCGCACCTCCCCACCGCCCGC
CGCCCGCCGCCCGCCGCCCGCAAAGCATGAGTGAGCCCGCTCTCTGCAGCTGCCCGGGGC
GCGAATGGCAGGCTGTTTCCGCGGAGTAAAAGGTGGCGCCGGTCAGTGGTCGTTTCCAAT
GACGGACATTAACCAGACTGTCAGATCCTGGGGAGTCGCGAGCCCCGAGTTTGGAGTTTT
TTCCCCCCACAACGTCACAGTCCGAACTGCAGAGGGAAAGGAAGGCGGCAGGAAGGCGAA
GCTCGGGCTCCGGCACGTAGTTGGGAAACTTGCGGGTCCTAGAAGTCGCCTCCCCGCCTT
GCCGGCCGCCCTTGCAGCCCCGAGCCGAGCAGCAAAGTGAGACATTGTGCGCCTGCCAGA
TCCGCCGGCCGCGGACCGGGGCTGCCTCGGAAACACAGAGGGGTCTTCTCTCGCCCTGCA
TATAATTAGCCTGCACACAAAGGGAGCAGCTGAATGGAGGTTGTCACTCTCTGGAAAAGG
ATTTCTGACCGAGCGCTTCCAATGGACATTCTCCAGTCTCTCTGGAAAGATTCTCGCTAA
TGGATTTCCTGCTGCTCGGTCTCTGTCTATACTGGCTGCTGAGGAGGCCCTCGGGGGTGG
TCTTGTGTCTGCTGGGGGCCTGCTTTCAGATGCTGCCCGCCGCCCCCAGCGGGTGCCCGC
AGCTGTGCCGGTGCGAGGGGCGGCTGCTGTACTGCGAGGCGCTCAACCTCACCGAGGCGC
CCCACAACCTGTCCGGCCTGCTGGGCTTGTCCCTGCGCTACAACAGCCTCTCGGAGCTGC
GCGCCGGCCAGTTCACGGGGTTAATGCAGCTCACGTGGCTCTATCTGGATCACAATCACA
TCTGCTCCGTGCAGGGGGACGCCTTTCAGAAACTGCGCCGAGTTAAGGAACTCACGCTGA
GTGCCTACCGGAGCTGCGGTGGCGTCTCCACACGCAACCATGAAGTTGAAGGACACAAAA
TCAAGGCCAAAGCAGTCAAGCTGTGGCAAATTTCAGACAAAGGGAATCAAAGTTGTGGGA
AAATGGAAGGAAATGGACAGATGGATGACTTGGTGTGCTTTGAGGAATTGACAGATTACC
AGTTGGTCTCCCCTGCCAAGAATCCCTCCAGTCTCTTCTCAAAGGAAGCACCCAAGAGAA
AGGCACAAGCTGTTTCAGAAGAAGAGGAGGAGGAGGAGGGAAAGTCTAGCTCACCAAAGA
AAAAGATCAAGTTGAAGAAAAGTAAAAATGTAGCAACTGAAGGAACCAGTACCCAGAAAG
AATTTGAAGTGAAAGATCCTGAGCTGGAGGCCCAGGGAGATGACATGGTTTGTGATGATC
CGGAGGCTGGGGAGATGACATCAGAAAACCTGGTCCAAACTGCTCCAAAAAAGAAGAAAA
ATAAAGGGAAAAAAGGGTTGGAGCCTTCTCAGAGCACTGCTGCCAAGGTGCCCAAAAAAG
CGAAGACATGGATTCCTGAAGTTCATGATCAGAAAGCAGATGTGTCAGCTTGGAAGGACC
TGTTTGTTCCCAGGCCGGTTCTCCGAGCACTCAGCTTTCTAGGCTTCTCTGCACCCACAC
CAATCCAAGCCCTGACCTTGGCACCTGCCATCCGTGACAAACTGGACATCCTTGGGGCTG
CTGAGACAGGAAGTGGGAAAACTCTTGCCTTTGCCATCCCAATGATTCATGCGGTGTTGC
AGTGGCAGAAGAGGAATGCTGCCCCTCCTCCAAGTAACACCGAAGCACCACCTGGAGAGA
CCAGAACTGAGGCCGGAGCTGAGACTAGATTACCAGGCAAGGCTGAAGCTGAGTCTGATG

CATTGCCTGACGATACTGTAATTGAGAGTGAAGCACTGCCCAGTGATATTGCAGCCGAGG
CCAGAGCCAAGACTGGAGGCACTGTCTCAGACCAGGCGTTGCTCTTTGAGTGACGATGAT
GCTGGTGAAGGGCCTTCTTCCCTGATCAGGGAGAAACCTGTTCCCAAACAGAATGGGAAT
GAAGAGGAAAATCTTTGATAAGAGCAGACTGGAAGTCTAAAACAGGAGTTGGATGA (SEQ ID NO:23)

TABLE 13B: The Amino Acid sequence of POLY12

MDFLLLGLCLYWLLRRPSGVVLCLLGACFQMLPAAPSGCPQLCRCEGRLLYCEALNLTEAPHNLSGLLGLSLRYNS
LSELRAGQFTGLMQLTWLYLDHNHICSVQGDAFQKLRRVKELTLSAYRSCGGVSTRNHEVEGHKIKAKAVKLWQIS
DKGNQSCGKMEGNGQMDDLVCFEELTDYQLVSPAKNPSSLFSKEAPKRKAQAVSEEEEEEEGKSSSPKKKIKLKKS
KNVATEGTSTQKEFEVKDPELEAQGDDMVCDDPEAGEMTSENLVQTAPKKKKNKGKKGLEPSQSTAAKVPKKAKTW
IPEVHDQKADVSAWKDLFVPRPVLRALSFLGFSAPTPIQALTLAPAIRDKLDILGAAETGSGKTLAFAIPMIHAVL
QWQKRNAAPPPSNTEAPPGETRTEAGAETRLPGKAEAESDALPDDTVIESEALPSDIAAEARAKTGGTVSDQALLF
E (SEQ ID NO:24)

In a search of sequence databases, it was found, for example, that POLY12 had 129 of 171 amino acid residues (75 %) identical to, and 135 of 171 residues (78%) similar to, the 172 amino acid (fragment) hematopoietic stem and progenitor cell 328 (HSPC328) protein from *Homo sapiens* (human) (ACC:AAF29006). SignalP, Psort and/or hydropathy suggest that POLY12 may be localized outside of the cell (Certainty= 0.6520) with a most likely cleavage site between positions 34 and 35 of SEQ ID NO.: 24. Since POLY12 is a member of the HSPC family, it is likely that this novel HSPC -like protein is available at the appropriate sub-cellular localization and hence accessible for the therapeutic uses described in this application.

A POLY12 nucleic acid is expressed in the following cells and tissues: kidney, skin, uterus, adrenal gland, placenta, hypothalamus, lymph node, fetal liver, bone marrow, fetal brain, fetal thymus, brain, HUVEC, salivary gland, testis, HuVec, CAEC, UtMVEC- myo, thyroid, PA-1, HEPG2, A204, HFDPC, stomach, trachea, SK-PN-DW, ovary tumor, breast carcinoma, CADMEC_LA, small intestine, hippocampus, Burkett's lymphoma, mammary gland, OVCAR-3, K-562, fetal lung, thalamus, spleen, and heart.

The expression pattern, and protein similarity information for the invention suggest that the human HSPC-like protein described in this invention may function a human HSPC-

like protein. Therefore, the nucleic acid and protein of the invention are useful in potential therapeutic applications implicated, for example but not limited to immune responses such as transplantation and inflammation and other diseases and disorders. The homology to antigenic secreted and membrane proteins suggests that antibodies directed against the novel genes may be useful in treatment and prevention of immune responses such as transplantation and inflammation, viral diseases, and other diseases and disorders. and other diseases and disorders.

The nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in immune responses such as transplantation and inflammation, viral diseases, and other diseases and disorders. For example, but not limited to, a cDNA encoding the human HSPC-like protein may be useful in gene therapy for hematopoietic disorders such as leukemia, and the human HSPC-like protein may be useful when administered to a subject in need thereof. By way of non-limiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from, for example, but not limited to, cancer, and other diseases and disorders. The novel nucleic acid encoding the human HSPC-like protein, and the human HSPC-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed. These materials are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in therapeutic or diagnostic methods.

POLY13

SULFOTRANSFERASE-LIKE PROTEINS AND NUCLEIC ACIDS

Sulfation is an important pathway in the biotransformation of steroid hormones such as estrogens. Human liver contains two different types of sulfotransferases, dehydroepiandrosterone (DHEA) sulfotransferase and phenol sulfotransferase. Estrogen preferring sulfotransferases are cytosolic proteins present in liver, intestine, and in kidney (at lower concentrations). Functionally, the enzyme is believed to control the level of the estrogen

receptor by sulfurylating free estradiol. It maximally sulfates beta-estradiol and estrone at concentrations of 20 nm, and dehydroepiandrosterone, pregnenolone, ethinylestradiol, equalenin, diethylstilbesterol, and 1-naphthol at significantly higher concentrations. However, cortisol, testosterone, and dopamine are not sulfated by the estrogen preferring sulfotransferases.

Cytosolic sulfotransferase (ST) enzymes catalyze the sulfate conjugation of many drugs, xenobiotic compounds, hormones, and neurotransmitters. There are a several STs with highly conserved regions among STs.

Estrone sulfate is the predominant form of estrogen found in the circulation in women and could thus serve as precursor for active estrogens in target tissues by removal of the sulfate group through the action of endogenous steroid sulfatase. A cDNA encoding human placental estrogen sulfotransferase was used as a probe for isolating a clone containing almost the whole genomic sequence. The gene contains nine short exons separated by eight introns in an expanse of approximately 7.7 kb. The first two exons, named exon 1a and exon 1b, are noncoding and correspond to the 5-prime untranslated sequences of human brain and human placental estrogen sulfotransferase cDNAs, respectively. Transfection of chloramphenicol acetyltransferase reporter gene vectors containing the 5-prime flanking sequence upstream from exon 1a and exon 1b into human adrenal adenocarcinoma cells indicated that both sequences possess promoter activity. The results were interpreted as indicating that human brain aryl sulfotransferase and placental estrogen sulfotransferase mRNA species are transcribed from a single gene by alternate exon 1a and exon 1b promoters, respectively. Using DNA from panels of human-rodent somatic cell hybrids and amplification of the gene by PCR, the placental estrogen sulfotransferase gene was assigned to chromosome 16, while liver estrogen sulfotransferase cDNA was mapped to 4q13.1 by fluorescence *in situ* hybridization, suggesting that these may be two separate genes. The liver STE gene spans approximately 20 kb and consists of eight exons, ranging in length from 95 to 181 bp. The locations of most exon-intron splice junctions within STE were identical to those found in a human phenol ST gene. The STM gene maps to chromosome 16p11.2. Indeed, STM is the same as the 'placental estrogen sulfotransferase' gene mapped to chromosome 16. The locations of five STE introns were conserved in the human DHEA-sulfotransferase gene,

which is located on chromosome 19. The Ste gene is located on mouse chromosome 5. All three known human STP genes, STP1, STP2, and STM, are located on 16p. STP1 is located approximately 45 kb 5-prime to STP2, and the 2 genes are aligned 'head-to-tail.' These 2 genes, in turn, are located approximately 100 kb telomeric to the gene for the monoamine-preferring sulfotransferase, STM. These three STP genes on 16p may be originated as a result of gene duplication events or gene duplication plus recombination. A mouse sulfotransferase gene, Stp, is located on mouse chromosome 7 in an area syntenic with human 16p.

A POLY13 nucleic acid was identified on chromosome 2 as described in Example 1. A nucleic acid of 921 nucleotides, POLY13 (designated CuraGen Acc. No. h_nh0443k08_A) encodes a novel Sulfotransferase-like protein as shown in TABLE 14. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 1-3 and ending with a TGA codon at nucleotides 916-918. A putative untranslated region upstream from the initiation codon and downstream from the termination codon is underlined in TABLE 14A, and the start and stop codons are in bold letters. The encoded protein having 305 amino acid residues is presented using the one-letter code in TABLE 14B.

TABLE 14A: The Nucleic Acid sequence of POLY13.

ATGGCGAAGATTGAGAAAAACGCTCCCACGATGGAAAAAAAGCCAGAACTGTTTAACATCATGGAAGTAG
ATGGAGTCCCTACGTTGATATTATCAAAAGAATGGTGGGAAAAAGTATGTAATTTCCAAGCCAAGCCTGA
TGATCTTATTCTGGCAACTTACCCAAAGTCAGGTACAACATGGATGCATGAAATTTTAGACATGATTCTA
AATGATGGTGATGTGGAGAAATGCAAAAGAGCCCAGACTCTAGATAGACACGCTTTCCTTGAACTGAAAT
TTCCCCATAAAGAAAAACCAGATTTGGAGTTCGTTCTTGAAATGTCCTCACCACAACTGATAAAAACACA
TCTCCCTTCACATCTGATTCCACCATCTATCTGGAAAGAAAACTGCAAGATTGTCTATGTGGCCAGAAAT
CCCAAGGATTGCCTGGTGTCCTACTACCACTTTCACAGGATGGCTTCCTTTATGCCTGATCCTCAGAACT
TAGAGGAATTTTATGAGAAATTCATGTCCGGAAAAGGTGAGTTCGGGTCCTGGTTTGACCATGTGAAAGG
ATGGTGGGCTGCAAAAGACATGCACCGGATCCTCTACCTCTTCTACGAGGATATTAAACAGAATCCAAAA
CATGAGATCCACAAGGTGTTGGAATTCTTGGAGAAAACTTGGTCAGGTGATGTTATAAACAAGATTGTCC
ACCATACCTCATTTGATGTAATGAAGGATAATCCCATGGCCAACCATACTGCGGTACCTGCTCACATATT
CAATCACTCCATCTCAAAATTTATGAGGAAAGGTGGGATGCCTGGAGACTGGAAGAACCACTTTACTGTG
GCTTTGAATGAGAACTTTGATAAGCATTATGAAAAGAAGATGGCAGGGTCCACACTGAACTTCTGCCTGG
AGATC**TGA**GAG (SEQ ID NO:25)

TABLE 14B: The Amino Acid sequence of POLY13.

MAKIEKNAPTMEKKPELFNIMEVDGVPTLILSKEWWEKVCNFQAKPDDLILATYPKSGTTWMHEILDMIL
NDGDVEKCKRAQTLDRHAFLELKFPHKEKPDLEFVLEMSSPQLIKTHLPSHLIPPSIWKENCKIVYVARN
PKDCLVSYYHFHRMASFMPDPQNLEEFYEKFMSGKGEFGSWFDHVKGWWAAKDMHRILYLFYEDIKQNPK
HEIHKVLEFLEKTWSGDVINKIVHHTSFDVMKDNPMANHTAVPAHIFNHSISKFMRKGGMPGDWKNHFTV
ALNENFDKHYEKKMAGSTLNFCLEI (SEQ ID NO :26)

In a search of sequence databases, it was found, for example, that the nucleic acid sequence has 647 of 920 bases (70%) identical to a *Mus musculus* Sulfotransferase mRNA (GENBANK-ID: AF033653|acc:AF033653). The full amino acid sequence of the protein of the invention was found to have 173 of 284 amino acid residues (60%) identical to, and 219 of 284 residues (77%) positive with, the 304 amino acid residue sulfotransferase protein from *Rattus norvegicus* (ptnr: SWISSPROT-ACC:P50237) (Table 14C).

TABLE 14C: BLASTX of POLY13 against N-Hydroxyarylamine Sulfotransferase (SEQ ID NO:40)

```
            Length = 304

 Score = 971 (341.8 bits), Expect = 3.5e-97, P = 3.5e-97
 Identities = 173/284 (60%), Positives = 219/284 (77%), Frame = +1

Query:    64 EVDGVPTLILSKEWWEKVCNFQAKPDDLILATYPKSGTTWMHEILDMILNDGDVEKCKRA 243
             ||+|+    |  + |+|+ ||||||||||++||| |+||||  ||+||| |||||+||+||
Sbjct:    22 EVNGILMSKLMSDNWDKIWNFQAKPDDLLIATYAKAGTTWTQEIVDMIQNDGDVQKCQRA 81

Query:   244 QTLDRHAFLELKFPHKEKPDLEFVLEMSSPQLIKTHLPSHLIPPSIWKENCKIVYVARNP 423
              | ||| |+|   |       |+   +| ||+ +||||| |++||| |||| ||+||||||
Sbjct:    82 NTYDRHPFIEWTLPSPLNSGLDLANKMPSPRTLKTHLPVHMLPPSFWKENSKIIYVARNA 141

Query:   424 KDCLVSYYHFHRMASFMPDPQNLEEFYEKFMSGKGEFGSWFDHVKGWWAAKDMHRILYLF 603
             ||||||||+| ||   +|||  | |+ |+| +||  +|||+|||||||  || |||||||
Sbjct:   142 KDCLVSYYYFSRMNKMLPDPGTLGEYIEQFKAGKVLWGSWYDHVKGWWDVKDQHRILYLF 201

Query:   604 YEDIKQNPKHEIHKVLEFLEKTWSGDVINKIVHHTSFDVMKDNPMANHTAVPAHIFNHSI 783
             |||+|++|| || |+ +|||| | +|+|||++|||||||+|||||+| +|+ | +|||
Sbjct:   202 YEDMKEDPKREIKKIAKFLEKDISEEVLNKIIYHTSFDVMKENPMANYTTLPSSIMDHSI 261

Query:   784 SKFMRKGGMPGDWKNHFTVALNENFDKHYEKKMAGSTLNFCLEI 915
             | ||||| ||||||||+|||| +|+||+ | +|||||  + |  ||
Sbjct:   262 SPFMRKG-MPGDWKNYFTVAQSEDFDEDYRRKMAGSNITFRTEI 304 (SEQ ID NO.: 40)
```

POLY13 also has significant homology to the proteins shown in the BLASTX data in Table 14D.

Table 14D. BLASTX alignments of POLY13

Sequences producing High-scoring Segment Pairs:	Reading Frame	High Score	Smallest Sum Prob. P(N)	N

patp:W40498	Human EST protein - Homo sapiens, 294 aa.+1	769	1.5e-75	1
patp:W44247	Human oestrogen sulphotransferase - Homo.+1	769	1.5e-75	1
patp:W23657	E6AP-binding protein cln25 - Homo sapien.+1	762	8.5e-75	1
patp:Y67294	Human STP2 (phenol sulphotransferase 2) .+1	744	6.8e-73	1
patp:Y45080	Wheat N-hydroxyarylamine sulphotransfera.+1	468	1.2e-43	1

The global sequence homology (as defined by FASTA alignment with the full length sequence of this protein) is 61% amino acid identity and 71% amino acid homology. In addition, this protein contains the following protein domains (as defined by Interpro) at the indicated amino acid positions: Sulfotransfer domain (IPR000863) at amino acid positions 24 to 293. PSORT analysis predicts the protein of the invention to be localized in the peroxisome with a certainty of 0.75. Based on the SIGNALP analysis, no signal peptide could be predicted for the protein of the invention.

The POLY13 nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in liver, intestine and kidney disorders including but not limited to primary biliary cirrhosis, primary sclerosing cholangitis, chronic active hepatitis and alcoholic cirrhosis, detoxification, ulcers, hyperthyroidism, developmental disorders, immune response, and/or other pathologies and disorders. For example, a cDNA encoding the Sulfotransferase-like protein may be useful in gene therapy in primary biliary cirrhosis, and the Sulfotransferase-like protein may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from liver, intestine and kidney disorders including but not limited to primary biliary cirrhosis, primary sclerosing cholangitis, chronic active hepatitis and alcoholic cirrhosis, detoxification, ulcers, hyperthyroidism, developmental disorders, various forms of immune response. The novel nucleic acid encoding Sulfotransferase-like protein, and the Sulfotransferase-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed. These materials are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in therapeutic or diagnostic methods.

POLY14-16

NOVEL SYNTAXIN-LIKE PROTEINS AND NUCLEIC ACIDS

Syntaxins belong to a family of proteins that appear to be involved in the docking vesicles with the plasma membrane during transmitter release. One of these proteins, designated syntaxin 1B (STX1B), is directly implicated in the process of calcium-dependent synaptic transmission in rat brain. The expression of this protein is transiently induced by long-term potentiation of synaptic responses in the rat hippocampus. The protein may play an important role in the excitatory pathway of synaptic transmission, which is known to be implicated in several neurologic diseases.The human STX1B gene was mapped to 16p11.2 by fluorescence *in situ* hybridization. The gene was found at a single locus. Chromosome rearrangements with breaks in 16p11 are observed in myxoid liposarcoma and in acute myeloid leukemia. A tumor that displays neuroendocrine properties, small cell lung cancer, has been observed in about 60% of patients with Lambert-Eaton myasthenic syndrome , an autoimmune disease of neurotransmission that is characterized by muscle weakness. Autoantibodies from these patients recognize the presynaptic N-type calcium channel and synaptotagmin, two proteins that are involved in synaptic transmission and interact with syntaxin.

Synaptic vesicles store neurotransmitters that are released during calcium-regulated exocytosis. The specificity of neurotransmitter release requires the localization of both synaptic vesicles and calcium channels to the presynaptic active zone. Syntaxins function in this vesicle fusion process. Syntaxins also serve as a substrate for botulinum neurotoxin type C, a metalloprotease that blocks exocytosis and has high affinity for a molecular complex that includes the alpha-latrotoxin receptor which produces explosive exocytosis.

By PCR analysis of human/rodent somatic cell hybrid panels and fluorescence *in situ* hybridization, the STX1A gene was mapped to chromosome 7q11.2. Syntaxin 1A is expressed in airway epithelial cells, and is not a neural-specific protein and syntaxin 1A regulates CFTR activity in airway epithelial cells.

Both syntaxin4 and VAMP2 are implicated in insulin regulation of glucose transporter-4 (GLUT4) trafficking in adipocytes as target (t) soluble N-ethylmaleimide-sensitive factor attachment protein receptors (SNARE) and vesicle (v)-SNARE proteins, respectively, which mediate fusion of GLUT4-containing vesicles with the plasma membrane. Synaptosome-associated 23-kDa protein (SNAP23) is a widely expressed isoform of SNAP25, the principal t-SNARE of neuronal cells, and colocalizes with syntaxin4 in the plasma membrane of 3T3-L1 adipocytes. In the present study, two SNAP23 mutants, SNAP23-DeltaC8 (amino acids 1 to 202) and SNAP23-DeltaC49 (amino acids 1 to 161), were generated to determine whether SNAP23 is required for insulin-induced translocation of GLUT4 to the plasma membrane in 3T3-L1 adipocytes. Wild-type SNAP23 (SNAP23-WT) promoted the interaction between syntaxin4 and VAMP2 both in vitro and in vivo. Although SNAP23-DeltaC49 bound to neither syntaxin4 nor VAMP2, the SNAP23-DeltaC8 mutant bound to syntaxin4 but not to VAMP2. In addition, although SNAP23-DeltaC8 bound to syntaxin4, it did not mediate the interaction between syntaxin4 and VAMP2. Moreover, overexpression of SNAP23-DeltaC8 in 3T3-L1 adipocytes by adenovirus-mediated gene transfer inhibited insulin-induced translocation of GLUT4 but not that of GLUT1. In contrast, overexpression of neither SNAP23-WT nor SNAP23-DeltaC49 in 3T3-L1 adipocytes affected the translocation of GLUT4 or GLUT1. Together, these results demonstrate that SNAP23 contributes to insulin-dependent trafficking of GLUT4 to the plasma membrane in 3T3-L1 adipocytes by mediating the interaction between t-SNARE (syntaxin4) and v-SNARE (VAMP2).

POLY14

A novel POLY14 nucleic acid was identified as described in Example 1. A POLY14 nucleic acid is found on chromosome 1. The novel nucleic acid of 893 nucleotides, POLY14 (designated CuraGen Acc. No. h_nh0778p17_A), encodes a novel Syntaxin-like protein as shown in TABLE 15. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 5-7 and ending with a TAA codon at nucleotides 887-889. A putative untranslated region upstream from the initiation codon and downstream from the termination codon is underlined in TABLE 15A, and the start and stop codons are in bold letters. The encoded protein having 294 amino acid residues is presented using the one-letter code in TABLE 15B.

TABLE 15A: The Nucleotide sequence of POLY14

GAAGATGAAAGACCGACTTCAAGAACTAAAGCAGAGAACAAAGGAAATTGAACTCTCTAGAGACAGTCAT
GTATCAACTACAGAAACAGAGGAACAAGGGGTGTTTCTACAGCAAGCTGTTATTTATGAAAGAGAGCCTG
TAGCTGAGAGACACCTACATGAAATCCAAAAACTACAGGAAAGTATTAACAATTTGGCAGATAATGTTCA
AAAATTTGGGCAGCAACAGAAAAGTCTGGTGGCTTCAATGAGAAGGTTTAGTCTACTTAAGAGAGAGTCT
ACCATTACAAAGGAGATAAAAATTCAGGCAGAATACATCAACAGAAGTTTGAATGATTTAGTTAAAGAAG
TTAAAAAGTCAGAGGTTGAAAATGGTCCATCTTCAGTGGTCACAAGGATACTTAAATCTCAGCATGCTGC
AATGTTCCGCCATTTTCAGCAAATCATGTTTATATACAATGACACAATAGCAGCAAAGCAAGAGAAGTGC
AAGACATTTATTTTACGTCAGCTTGAAGTTGCTGGAAAAGAGATGTCTGAAGAAGATGTAAATGATATGC
TTCATCAAGGAAAATGGGAAGTTTTTAATGAAAGCTTACTTACAGAAATCAATATCACTAAAGCACAACT
TTCAGAGATTGAACAGAGACACAAGGAACTTGTTAATTTGGAGAACCAAATAAAGGATTTAAGGGATCTT
TTCATTCAGATATCTCTTTTAGTAGAGGAACAAGGAGAGAGCATCAACAATATTGAAATGACAGTGAATA
GTACAAAAGAGTATGTTAACAATACTAAAGAGAAATTTGGACTAGCTGTAAAATACAAAAAAAGAAATCC
TTGCAGAGTACTGTGTTGTTGGTGCTGTCCATGCTGTAGCTCAAAATAAAGAA (SEQ ID NO:27)

TABLE 15B: The Amino Acid sequence of POLY14

MKDRLQELKQRTKEIELSRDSHVSTTETEEQGVFLQQAVIYEREPVAERHLHEIQKLQESINNLADNVQK
FGQQQKSLVASMRRFSLLKRESTITKEIKIQAEYINRSLNDLVKEVKKSEVENGPSSVVTRILKSQHAAM
FRHFQQIMFIYNDTIAAKQEKCKTFILRQLEVAGKEMSEEDVNDMLHQGKWEVFNESLLTEINITKAQLS
EIEQRHKELVNLENQIKDLRDLFIQISLLVEEQGESINNIEMTVNSTKEYVNNTKEKFGLAVKYKKRNPC
RVLCCWCCPCCSSK (SEQ ID NO :28)

In a search of sequence databases, it was found, for example, that the nucleic acid sequence has 363 of 622 bases (58%) identical to a Caenorhabditis elegans Syntaxin mRNA (GENBANK-ID: AB008842|acc:AB008842). The full amino acid sequence of the protein of the invention was found to have 108 of 290 amino acid residues (37%) identical to, and 184 of 290 residues (63%) positive with, the 287 amino acid residue SYNTAXIN 11 protein from Homo sapiens (ptnr: SWISSNEW-ACC:O75558) (Table 15C).

TABLE 15C: BLASTX of POLY14 against Syntaxin 11 (SEQ ID NO:41)

```
Length = 287

Score = 553 (194.7 bits), Expect = 6.9e-53, P = 6.9e-53
 Identities = 108/290 (37%), Positives = 184/290 (63%), Frame = +2

Query:     5 MKDRLQELKQRTKEIELSRDSHVSTTETEEQGVFLQQAVIYEREPVAERHLHEIQKLQES 184
             |||||  ||      ++||+       + +++     + +++| + + |    +|+ +|+
Sbjct:     1 MKDRLAEL------LDLSKQYDQQFPDGDDEFDSPHEDIVFETDHILESLYRDIRDIQDE 54

Query:   185 INNLADNVQKFGQQQKSLVASMRRFSLLKRES-TITKEIKIQAEYINRSLNDLVKEVKKS 361
               |  +|++ |+|    + |||| | +||++ +| | || + | |+  |  + +  + +
Sbjct:    55 NQLLVADVKRLGKQNARFLTSMRRLSSIKRDTNSIAKAIKARGEVIHCKLRAMKELSEAA 114

Query:   362 EVENGPSSVVTRILKSQHAAMFRHFQQIMFIYNDTIAAKQEKCKTFILRQLEVAGKEMSE 541
             | ++|| | | || ++|+ |+   ||+ |  ||     +++ ||  | ||||+ |||+|
Sbjct:   115 EAQHGPHSAVARISRAQYNALTLTFQRAMHDYNQAEMKQRDNCKIRIQRQLEIMGKEVSG 174
```

```
Query:   542 EDVNDMLHQGKWEVFNESLLTEINITKAQLSEIEQRHKELVNLENQIKDLRDLFIQISLL 721
             + + ||  ||||+||+|+|| ++   +| |+||| ||+||+ ||++|+|+ +||+|+++|
Sbjct:   175 DQIEDMFEQGKWDVFSENLLADVKGARAALNEIESRHRELLRLESRIRDVHELFLQMAVL 234

Query:   722 VEEQGESINNIEMTVNSTKEYVNNTKEKFGLAVKYKKRNPCRVLCCWCCPC 874
             ||+| +++| ||+ |  | +|   | +  ||+|+++|||| |||+||||
Sbjct:   235 VEKQADTLNVIELNVQKTVDYTGQAKAQVRKAVQYEEKNPCRTLCCFCCPC 285 (SEQ ID NO.41)
```

POLY14 also has significant homology to the proteins shown in the BLASTX data in Table 15D.

Table 15D. BLASTX alignments of POLY14

Sequences producing High-scoring Segment Pairs:	Reading Frame	High Score	Smallest Sum Prob. P(N)	N
patp:R44916 Rat post-synaptic NMDA receptor GR33 - Rat.	+2	381	2.0e-34	1
patp:W43419 Rat syntaxin 1B protein - Rattus sp, 288 aa	+2	381	2.0e-34	1
patp:R96421 Rat syntaxin 1A - Rattus rattus, 288 aa.	+2	379	3.3e-34	1
patp:W30105 Rat syntaxin - Rattus sp, 288 aa.	+2	379	3.3e-34	1
patp:W24927 Rat syntaxin 1A - Rattus rattus, 288 aa.	+2	379	3.3e-34	1
patp:B12822 Rat syntaxin 1A amino acid sequence - Ratt.	+2	379	3.3e-34	1

The global sequence homology (as defined by FASTA alignment with the full length sequence of this protein) is 51% amino acid identity and 37% amino acid homology. In addition, this protein contains the following protein domain (as defined by Interpro) at the indicated amino acid positions: syntaxin family (IPR000017) at amino acid positions 1-292. PSORT analysis predicts the protein of the invention to be localized in the plasma membrane with a certainty of 0.6. Based on the SIGNALP analysis, no signal peptide could be predicted for the protein of the invention.

POLY15

In the present invention, the target sequence identified previously, POLY14, was subjected to the exon linking process as described in Example 6. The novel nucleic acid of 892 nucleotides, POLY15 (designated CuraGen Acc. No. h_nh0778p17_A1), encodes a novel

Syntaxin-like protein as shown in TABLE 16. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 4-6 and ending with a TAA codon at nucleotides 887-889. A putative untranslated region upstream from the initiation codon and downstream from the termination codon is underlined in TABLE 16A, and the start and stop codons are in bold letters. The encoded protein having 294 amino acid residues is presented using the one-letter code in TABLE 16B. The molecular weight of POLY15 is 3.4324 kDa.

TABLE 16A: The Nucleotide sequence of POLY15

AAGATGAAAGACCGACTTCAAGAACTAAAGCAGAGAACAAAGGAAATTGAACTCTCTAGAGACAGTCATGTATCAA
CTACAGAAACAGAGGAACAAGGGGTGTTTCTACAGCAAGCTGTTATTTATGAAAGAGAGCCTGTAGCTGAGAGACA
CCTACATGAAATCCAAAAACTACAGGAAAGTATTAACAATTTGGCAGATAATGTTCAAAAATTTGGGCAGCAACAG
AAAAGTCTGGTGGCTTCAATGAGAAGGTTTAGTCTACTTAAGAGAGAGTCTACCATTACAAAGGAGATAAAAATTC
AGGCAGAATACATCAACAGAAGTTTGAATGATTTAGTTAAAGAAGTTAAAAAGTCAGAGGTTGAAAATGGTCCATC
TTCAGTGGTCACAAGGATACTTAAATCTCAGCATGCTGCAATGTTCCGCCATTTTCAGCAAATCATGTTTATATAC
AATGACACAATAGCAGCAAAGCAAGAGAAGTGCAAGACATTTATTTTACGTCAGCTTGAAGTTGCTGGAAAAGAGA
TGTCTGAAGAAGATGTAAATGATATGCTTCATCAAGGAAAATGGGAAGTTTTTAATGAAAGCTTACTTACAGAAAT
CAATATCACTAAAGCACAACTTTCAGAGATTGAACAGAGACACAAGGAACTTGTTAATTTGGAGAACCAAATAAAG
GATTTAAGGGATCTTTTCATTCAGATATCTCTTTTAGTAGAGGAACAAGGAGAGAGCATCAACAATATTGAAATGA
CAGTGAATAGTACAAAAGAGTATGTTAACAATACTAAAGAGAAATTTGGACTAGCTGTAAAATACAAAAAAAGAAA
TCCTTGCAGAGTACTGTGTTGTTGGTGCTGTCCATGCTGTAGCTCAAAATAAAGAA (SEQ ID NO:29)

TABLE 16B: The Amino Acid sequence of POLY15

MKDRLQELKQRTKEIELSRDSHVSTTETEEQGVFLQQAVIYEREPVAERHLHEIQKLQESINNLADNVQKFGQQQK
SLVASMRRFSLLKRESTITKEIKIQAEYINRSLNDLVKEVKKSEVENGPSSVVTRILKSQHAAMFRHFQQIMFIYN
DTIAAKQEKCKTFILRQLEVAGKEMSEEDVNDMLHQGKWEVFNESLLTEINITKAQLSEIEQRHKELVNLENQIKD
LRDLFIQISLLVEEQGESINNIEMTVNSTKEYVNNTKEKFGLAVKYKKRNPCRVLCCWCCPCCSSK (SEQ ID NO :30)

The full amino acid sequence of the protein of the invention was found to have 108 of 290 amino acid residues (37%) identical to, and 184 of 290 residues (63%) positive with, the 287 amino acid residue SYNTAXIN 11 protein from Homo sapiens (ptnr: SWISSNEW-ACC:O75558) (Table 16C).

TABLE 16C: BLASTX of POLY15 against Syntaxin 11 (SEQ ID NO:42)

```
            Length = 287

  Score = 553 (194.7 bits), Expect = 6.9e-53, P = 6.9e-53
 Identities = 108/290 (37%), Positives = 184/290 (63%), Frame = +2

Query:     5 MKDRLQELKQRTKEIELSRDSHVSTTETEEQGVFLQQAVIYEREPVAERHLHEIQKLQES 184
             ||||| ||      ++||+       + +++     + +++| + + |    +|+ +|+
```

```
Sbjct:     1 MKDRLAEL------LDLSKQYDQQFPDGDDEFDSPHEDIVFETDHILESLYRDIRDIQDE 54

Query:   185 INNLADNVQKFGQQQKSLVASMRRFSLLKRES-TITKEIKIQAEYINRSLNDLVKEVKKS 361
               |  +|++ |+|    + |||| | +||++ +| | || + | |+  |  + +  + +
Sbjct:    55 NQLLVADVKRLGKQNARFLTSMRRLSSIKRDTNSIAKAIKARGEVIHCKLRAMKELSEAA 114

Query:   362 EVENGPSSVVTRILKSQHAAMFRHFQQIMFIYNDTIAAKQEKCKTFILRQLEVAGKEMSE 541
               | ++|| | | || ++|+ |+   ||+ |  ||     +++ ||  | ||||+ |||+|
Sbjct:    115 EAQHGPHSAVARISRAQYNALTLTFQRAMHDYNQAEMKQRDNCKIRIQRQLEIMGKEVSG 174

Query:   542 EDVNDMLHQGKWEVFNESLLTEINITKAQLSEIEQRHKELVNLENQIKDLRDLFIQISLL 721
               + + ||  ||||+||+|+|| ++   +| |+||| ||+||+ ||++|+|+ +||+|+++|
Sbjct:    175 DQIEDMFEQGKWDVFSENLLADVKGARAALNEIESRHRELLRLESRIRDVHELFLQMAVL 234

Query:   722 VEEQGESINNIEMTVNSTKEYVNNTKEKFGLAVKYKKRNPCRVLCCWCCPC 874
               ||+| +++| ||+ |  | +|    | +   ||+|+++|||| |||+||||
Sbjct:    235 VEKQADTLNVIELNVQKTVDYTGQAKAQVRKAVQYEEKNPCRTLCCFCCPC 285 (SEQ ID NO. 42)
```

The global sequence homology (as defined by FASTA alignment with the full length sequence of this protein) is 51% amino acid identity and 37% amino acid homology. In addition, this protein contains the following protein domain (as defined by Interpro) at the indicated amino acid positions: syntaxin family (IPR000017) at amino acid positions 1-292.

PSORT analysis predicts the protein of the invention to be localized in the cytoplasm with a certainty of 0.6500. Based on the SIGNALP analysis, no N-terminal signal peptide could be predicted for the protein of the invention.

POLY16

In the present invention, the target sequence identified previously, POLY14, was subjected to the exon linking process as described in Example 6. The novel nucleic acid of 892 nucleotides, POLY16 (designated CuraGen Acc. No. CG55655-02), encodes a novel Syntaxin-like protein as shown in TABLE 17. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 4-6 and ending with a TAA codon at nucleotides 887-889. A putative untranslated region upstream from the initiation codon and downstream from the termination codon is underlined in TABLE 17A, and the start and stop codons are in bold letters. The encoded protein having 294 amino acid residues is presented using the one-letter code in TABLE 17B. The molecular weight of POLY16 is 3.4324 kDa.

TABLE 17A: The Nucleotide sequence of POLY16

AAGATGAAAGACCGACTTCAAGAACTAAAGCAGAGAACAAAGGAAATTGAACTCTCTAGAGACAGTCATGTATCAA
CTACAGAAACAGAGGAACAAGGGGTGTTTCTACAGCAAGCTGTTATTTATGAAAGAGAGCCTGTAGCTGAGAGACA
CCTACATGAAATCCAAAAACTACAGGAAAGTATTAACAATTTGGCAGATAATGTTCAAAAATTTGGGCAGCAACAG
AAAAGTCTGGTGGCTTCAATGAGAAGGTTTAGTCTACTTAAGAGAGAGTCTACCATTACAAAGGAGATAAAAATTC
AGGCAGAATACATCAACAGAAGTTTGAATGATTTAGTTAAAGAAGTTAAAAAGTCAGAGGTTGAAAATGGTCCATC
TTCAGTGGTCACAAGGATACTTAAATCTCAGCATGCTGCAATGTTCCGCCATTTTCAGCAAATCATGTTTATATAC
AATGACACAATAGCAGCAAAGCAAGAGAAGTGCAAGACATTTATTTTACGTCAGCTTGAAGTTGCTGGAAAAGAGA
TGTCTGAAGAAGATGTAAATGATATGCTTCATCAAGGAAAATGGGAAGTTTTTAATGAAAGCTTACTTACAGAAAT
CAATATCACTAAAGCACAACTTTCAGAGATTGAACAGAGACACAAGGAACTTGTTAATTTGGAGAACCAAATAAAG
GATTTAAGGGATCTTTTCATTCAGATATCTCTTTTAGTAGAGGAACAAGGAGAGAGCATCAACAATATTGAAATGA
CAGTGAATAGTACAAAAGAGTATGTTAACAATACTAAAGAGAAATTTGGACTAGCTGTAAAATACAAAAAAAGAAA
TCCTTGCAGAGTACTGTGTTGTTGGTGCTGTCCATGCTGTAGCTCAAAATAAAGAA (SEQ ID NO:29)

TABLE 17B: The Amino Acid sequence of POLY16

MKDRLQELKQRTKEIELSRDSHVSTTETEEQGVFLQQAVIYEREPVAERHLHEIQKLQESINNLADNVQKFGQQQK
SLVASMRRFSLLKRESTITKEIKIQAEYINRSLNDLVKEVKKSEVENGPSSVVTRILKSQHAAMFRHFQQIMFIYN
DTIAAKQEKCKTFILRQLEVAGKEMSEEDVNDMLHQGKWEVFNESLLTEINITKAQLSEIEQRHKELVNLENQIKD
LRDLFIQISLLVEEQGESINNIEMTVNSTKEYVNNTKEKFGLAVKYKKRNPCRVLCCWCCPCCSSK (SEQ ID
NO :30)

The full amino acid sequence of the protein of the invention was found to have 108 of 290 amino acid residues (37%) identical to, and 184 of 290 residues (63%) positive with, the 287 amino acid residue SYNTAXIN 11 protein from Homo sapiens (ptnr: SWISSNEW-ACC:O75558). The global sequence homology (as defined by FASTA alignment with the full length sequence of this protein) is 51% amino acid identity and 37% amino acid homology. In addition, this protein contains the following protein domain (as defined by Interpro) at the indicated amino acid positions: syntaxin family (IPR000017) at amino acid positions 1-292.

PSORT analysis predicts the protein of the invention to be localized in the cytoplasm with a certainty of 0.6500. Based on the SIGNALP analysis, no N-terminal signal peptide could be predicted for POLY16.

This Syntaxin-like protein may function as a member of a "Syntaxin family". Therefore, the POLY14-16 novel nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders such as various forms of cancers, neurologic disorders, autoimmune disease, CFTR, Lambert-Eaton myasthenic syndrome, small cell lung cancer, myxoid liposarcoma and in acute myeloid leukemia, Type I and II diabetes, obesity, skin disorders, degenerative disorders

affecting epithelial-derived tissues and/or other pathologies and disorders. For example, a cDNA encoding the Syntaxin-like protein may be useful in gene therapy, and the Syntaxin-like protein may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from various forms of cancers, neurologic disorders, autoimmune disease, CFTR, Lambert-Eaton myasthenic syndrome, small cell lung cancer, myxoid liposarcoma and in acute myeloid leukemia, Type I and II diabetes, obesity, skin disorders, and various degenerative disorders affecting epithelial-derived tissues. The novel nucleic acid encoding Syntaxin-like protein, and the Syntaxin-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed. These materials are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in therapeutic or diagnostic methods.

POLY17

PROHIBITIN-LIKE PROTEINS AND NUCLEIC ACIDS

Prohibitin (PHB) is a 30-kD intracellular, antiproliferative protein. The gene was mapped to chromosome 17 by analysis of human-mouse somatic cell hybrid cell lines using a genomic fragment of human prohibitin DNA isolated from a library using the rat prohibitin cDNA clone. The PHB gene was located in the 17q11.2-q23 region where a gene responsible for hereditary breast cancer is localized. The human homolog of the rat prohibitin gene mapped to 17q12-q21 by *in situ* hybridization. The human prohibitin gene family consists of one functional PHB gene on 17q21 and four processed pseudogenes, each on a different chromosome: PHBP1 on 6q25, PHBP2 on 11p11.2, PHBP3 on 1p31.3, and PHBP4 on 2q21.

DNA sequence analysis of two exons in this gene in 23 sporadic breast cancers, which showed loss of heterozygosity on the long arm of chromosome 17 (17q) or developed in patients 35 years old or younger, identified four cases of somatic mutation; two of these were missense mutations; one showed a two-base deletion resulting in truncation of the gene

product due to a frame shift; the other had a C to T transition in an intron adjacent to an intron-exon boundary. These results suggest that this gene may be a tumor suppressor gene and is associated with tumor development and/or progression of at least some breast cancers. Mutations in the PHB gene were not detected in other forms of tumors, namely, those of ovary, liver, and lung.

The retinoblastoma tumor suppressor protein and its family members, p107 and p130, are major regulators of the mammalian cell cycle. They exert their growth suppressive effects at least in part by binding the E2F family of transcription factors and inhibiting their transcriptional activity. Agents that disrupt the interaction between Rb family proteins and E2F promote cell proliferation. Prohibitin physically interacts with all three Rb family proteins in vitro and in vivo, and was very effective in repressing E2F-mediated transcription. Prohibitin could inhibit the activity of E2Fs 1, 2, 3, 4 and 5, but could not affect the activity of promoters lacking an E2F site. Prohibitin-mediated repression of E2F could not be reversed by adenovirus E1A protein. A prohibitin mutant that could not bind to Rb was impaired in its ability to repress E2F activity and inhibit cell proliferation. Prohibitin may be a novel regulator of E2F activity that responds to specific signaling cascades.

A POLY17 nucleic acid was identified as described in Example 1. A POLY17 nucleic acid of 967 nucleotides (designated CuraGen Acc. No. GM_11817402_A) encoding a novel Prohibitin -like protein is shown in TABLE 18. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 75-77 and ending with a TGA codon at nucleotides 888-890. A putative untranslated region upstream from the initiation codon and downstream from the termination codon is underlined in TABLE 18A, and the start and stop codons are in bold letters. The encoded protein having 271 amino acid residues is presented using the one-letter code in TABLE 18B.

TABLE 18A: The Nucleic Acid sequence of POLY17.

TCAGAAATCAATGATAAAGGGACGGAATTCATGTGGGGGGTTGGAGTGGACGCAGGCGTGAGTGGGTCCAGCA
G**ATG**GAAACACAGCTGCCAAGTCTGCCCCTGTCCTTAGCTTCTGCAGGAGGTGTGGGGAACTCTGCCTTCTAC
AATGTGATGCTGCACAGAGAGCTGTCTGTCATCTTCGACCAATTCCATGGCATTCAGGACACTGTGATAGGGG
AAGGAACGCACTTTCTCATCCCATGGGAAAAGAAACCAATTATTTTTGACTGCTGCTCTCGACCACATTATGC
ACCAATCATCACTGTGAGCAAAGATTGTCACCATGTCACCATCACACTGGGCGTCCTCTTCCCGCCTTGTTGC

TGGCCAGGTCCTTGCATCTTCCAATTACTGGAGAAGCCAATGAAGAATGTGCTGCCATCCATCACTGCGGAGC
TCCTCAAGCTGGGGGCGGCTCAGGCTGACGCTGGAGAACTGATCACGCAGGGAGAGCTGGGCTCCAGACAGGT
GAGCGAGCAATTAACTGAGCAAGCAGCAACCTTTGGGTTCCTCCTGGATGCTGTGACCTTGGATCTGACCTTC
GGGAAGGAATTTGCAGAAGCAGTGGAACCAAAGGAGGTGGCTCAGCAGGAAGAAGAGAGGGCCAGATCTGTGG
TGGCAAGGGCTGAGCAGCAGAAGACGGCGGCCATCATCTCTGCCGAGGGCGACTCCAAGGCCACGGAGTTCAT
CGCCAGCTCAGTGGCCACCGCAGGTGACGGCCTGATCAAGGCCCACAAGCTGGAACCATGGAGGACACTGGCC
CTCCAGCTCTCAGAACTCATCCACCTCATCCACCTGCCCGTGGGGACATCTGTGCTCCTCCAGCTGCCCCAGC
GCAGGCCGCCC**TGA**CCTGCACCTCCTCCAGCCAACTGGGCCACAGCACCAATGACTTTTACTACCGCCTTCCT
TCTGTCCCCACTCCAGAA (SEQ ID NO:33)

TABLE 18B: The Amino Acid sequence of POLY17.

METQLPSLPLSLASAGGVGNSAFYNVMLHRELSVIFDQFHGIQDTVIGEGTHFLIPWEKKPII
FDCCSRPHYAPIITVSKDCHHVTITLGVLFPPCCWPGPCIFQLLEKPMKNVLPSITAELLKLG
AAQADAGELITQGELGSRQVSEQLTEQAATFGFLLDAVTLDLTFGKEFAEAVEPKEVAQQEEE
RARSVVARAEQQKTAAIISAEGDSKATEFIASSVATAGDGLIKAHKLEPWRTLALQLSELIHL
IHLPVGTSVLLQLPQRRPP (SEQ ID NO:34)

In a search of sequence databases, it was found, for example, that the nucleic acid sequence has 737 of 936 bases (78%) identical to *Homo sapiens* Prohibitin mRNA (GENBANK-ID: S85655). The full amino acid sequence of the protein of the invention was found to have 168 of 259 amino acid residues (64%) identical to, and 194 of 259 residues (74 %) positive with, the 272 amino acid residue Prohibitin protein from *Homo sapiens* (ptnr:SPTREMBL-ACC:P35232).

POLY17 also has high homology to the proteins in the BLAST data shown in Table 18C.

Table 18C. BLASTP results for POLY17

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
patp:B43874	Human cancer associated protein sequence	279	168/259 (64%)	194/259 (74%)	1.6e-76
patp:W54352	Heat shock 27 kD protein and prohibitin (admixture) - Homo sapiens	471	173/281 (61%)	204/281 (72%)	2.3e-74
patp:R42215	Human prohibitin	272	168/259 (64%)	194/259 (74%)	6.0e-74

patp:R13466	Prohibitin - Rattus rattus	272	167/259 (64%)	194/259 (74%)	1.2e-73

POLY17 also has high homology to the amino acid sequences presented in the BLASTX data shown in Table 18D.

Table 18D. BLASTX alignments of POLY17

Sequences producing High-scoring Segment Pairs:	Reading Frame	High Score	Smallest Sum Prob. P(N)	N
patp:R13467 Cc protein - Drosophila, 276 aa.	+3	613	5.2e-59	1
patp:B65735 Prohibitin-related protein #1 - Pinus radi.	+3	473	3.6e-44	1

The global sequence homology (as defined by FASTA alignment with the full length sequence of this protein) is 68% amino acid identity and 63% amino acid homology. In addition, this protein contains the following protein domains (as defined by Interpro) at the indicated amino acid positions: SPFH domain / Band 7 family (IPR001107) at amino acid positions from 8 to 202.

In a search of CuraGen's proprietary human expressed sequence assembly database, assembly 11817402 (309 nucleotides) was/were identified as having >95% homology to this predicted gene sequence. This database is composed of the expressed sequences (as derived from isolated mRNA) from more than 96 different tissues. The mRNA is converted to cDNA and then sequenced. These expressed DNA sequences are then pooled in a database and those exhibiting a defined level of homology are combined into a single assembly with a common consensus sequence. The consensus sequence is representative of all member components. Since the nucleic acid of the described invention has >95% sequence identity with the CuraGen assembly, the nucleic acid of the invention represents an expressed gene sequence. This DNA assembly has one component and was found by CuraGen to be expressed in the endocrine system, for example in the thyroid.

PSORT analysis predicts the protein of the invention to be localized in the cytoplasm with a certainty of 0.4500. Using the SIGNALP analysis, it is predicted that the protein of the invention has a signal peptide with most likely cleavage site between pos. 19 and 20 of SEQ ID NO.: 34.

POLY17 is a new member of the prohibitin-like family of proteins, and is therefor useful as a marker to detect binding proteins of the probibitin-like protein family. POLY17 is also useful to detect tissues of the endocrine system, *e.g.* the thyroid, and activated B-cells, *e.g.* PMA-treated chronic leukemic B-cells. The above defined information for POLY 17 suggests that this Prohibitin -like protein may function as a member of a "Prohibitin family". Therefore, the novel nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders such as breast and ovarian cancer, tumor suppression, senescence, growth regulation, modulation of apotosis, reproductive control and associated disorders of reproduction, endometrial hyperplasia and adenocarcinoma, and/or other pathologies and disorders. For example, a cDNA encoding the Prohibitin -like protein may be useful in gene therapy for leukemia, and the Prohibitin -like protein may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from cancers including but not limited to breast and ovarian cancer, tumor suppression, senescence, growth regulation, modulation of apotosis, reproductive control and associated disorders of reproduction, endometrial hyperplasia and adenocarcinoma. The novel nucleic acid encoding Prohibitin -like protein, and the Prohibitin -like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed. These materials are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in therapeutic or diagnostic methods.

POLYX Nucleic Acids

The novel nucleic acids of the invention include those that encode a POLYX or POLYX-like protein, or biologically-active portions thereof. The nucleic acids include nucleic acids encoding polypeptides that include the amino acid sequence of one or more of SEQ ID NO:2*n* (wherein n = 1 to 17). The encoded polypeptides can thus include, *e.g.*, the amino acid sequences of SEQ ID NO:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32 and/or 3414, 16, 18, 20, 22, 24, 26, 28, 30, 32 and/or 34.

In some embodiments, a nucleic acid encoding a polypeptide having the amino acid sequence of one or more of SEQ ID NO:2*n* (wherein n = 1 to 17) includes the nucleic acid sequence of any of SEQ ID NO:2*n*-1 (wherein n = 1 to 17), or a fragment thereof, and can thus include, *e.g.*, the nucleic acid sequences of SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31 and/or 33. Additionally, the invention includes mutant or variant nucleic acids of any of SEQ ID NO:2*n*-1 (wherein n = 1 to 17), or a fragment thereof, any of whose bases may be changed from the disclosed sequence while still encoding a protein that maintains its POLYX-like biological activities and physiological functions. The invention further includes the complement of the nucleic acid sequence of any of SEQ ID NO:2*n*-1 (wherein n = 1 to 17), including fragments, derivatives, analogs and homologs thereof. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications.

Also included are nucleic acid fragments sufficient for use as hybridization probes to identify POLYX-encoding nucleic acids (*e.g.*, POLYX mRNA) and fragments for use as polymerase chain reaction (PCR) primers for the amplification or mutation of POLYX nucleic acid molecules. As used herein, the term "nucleic acid molecule" is intended to include DNA molecules (*e.g.*, cDNA or genomic DNA), RNA molecules (*e.g.*, mRNA), analogs of the DNA or RNA generated using nucleotide analogs, and derivatives, fragments, and homologs thereof. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA.

As utilized herein, the term "probes" refer to nucleic acid sequences of variable length, preferably between at least about 10 nucleotides (nt), 100 nt, or as many as about, *e.g.*, 6,000 nt, depending upon the specific use. Probes are used in the detection of identical, similar, or complementary nucleic acid sequences. Longer length probes are usually obtained from a natural or recombinant source, are highly specific and much slower to hybridize than oligomers. Probes may be single- or double-stranded, and may also be designed to have specificity in PCR, membrane-based hybridization technologies, or ELISA-like technologies.

As utilized herein, the term "isolated" nucleic acid molecule is a nucleic acid that is separated from other nucleic acid molecules that are present in the natural source of the nucleic acid. Examples of isolated nucleic acid molecules include, but are not limited to, recombinant DNA molecules contained in a vector, recombinant DNA molecules maintained in a heterologous host cell, partially or substantially purified nucleic acid molecules, and synthetic DNA or RNA molecules. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (*i.e.*, sequences located at the 5'- and 3'-termini of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated POLYX nucleic acid molecule can contain less than approximately 50 kb, 25 kb, 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material or culture medium when produced by recombinant techniques, or of chemical precursors or other chemicals when chemically synthesized.

As used herein, a "mature" form of a polypeptide or protein disclosed in the present invention is the product of a naturally occurring polypeptide or precursor form or proprotein. The naturally occurring polypeptide, precursor or proprotein includes, by way of nonlimiting example, the full length gene product, encoded by the corresponding gene. Alternatively, it may be defined as the polypeptide, precursor or proprotein encoded by an open reading frame described herein. The product "mature" form arises, again by way of nonlimiting example, as a result of one or more naturally occurring processing steps as they may take place within the cell, or host cell, in which the gene product arises. Examples of such processing steps leading

to a "mature" form of a polypeptide or protein include the cleavage of the N-terminal methionine residue encoded by the initiation codon of an open reading frame, or the proteolytic cleavage of a signal peptide or leader sequence. Thus a mature form arising from a precursor polypeptide or protein that has residues 1 to N, where residue 1 is the N-terminal methionine, would have residues 2 through N remaining after removal of the N-terminal methionine. Alternatively, a mature form arising from a precursor polypeptide or protein having residues 1 to N, in which an N-terminal signal sequence from residue 1 to residue M is cleaved, would have the residues from residue M+1 to residue N remaining. Further as used herein, a "mature" form of a polypeptide or protein may arise from a step of post-translational modification other than a proteolytic cleavage event. Such additional processes include, by way of non-limiting example, glycosylation, myristoylation or phosphorylation. In general, a mature polypeptide or protein may result from the operation of only one of these processes, or a combination of any of them.

A nucleic acid molecule of the invention, *e.g.*, a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:2*n*-1 (wherein n = 1 to 17), or a complement of any of these nucleotide sequences, can be isolated using standard molecular biology techniques and the sequence information provided herein. Using all or a portion of the nucleic acid sequence of any of SEQ ID NO:2*n*-1 (wherein *n*=1 to 17) as a hybridization probe, POLYX nucleic acid sequences can be isolated using standard hybridization and cloning techniques (*e.g.*, as described in Sambrook *et al.*, eds., MOLECULAR CLONING: A LABORATORY MANUAL 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989; and Ausubel, *et al.*, eds., CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY, 1993.)

A nucleic acid of the invention can be amplified using cDNA, mRNA or alternatively, genomic DNA, as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to POLYX nucleotide sequences can be prepared by standard synthetic techniques, *e.g.*, using an automated DNA synthesizer.

As used herein, the term "oligonucleotide" refers to a series of linked nucleotide residues, which oligonucleotide has a sufficient number of nucleotide bases to be used in a PCR reaction. A short oligonucleotide sequence may be based on, or designed from, a genomic or cDNA sequence and is used to amplify, confirm, or reveal the presence of an identical, similar or complementary DNA or RNA in a particular cell or tissue. Oligonucleotides comprise portions of a nucleic acid sequence having about 10 nt, 50 nt, or 100 nt in length, preferably about 15 nt to 30 nt in length. In one embodiment, an oligonucleotide comprising a nucleic acid molecule less than 100 nt in length would further comprise at lease 6 contiguous nucleotides of any of SEQ ID NO:2*n*-1 (wherein n = 1 to 17), or a complement thereof. Oligonucleotides may be chemically synthesized and may also be used as probes.

In another embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule that is a complement of the nucleotide sequence shown in any of SEQ ID NO:2*n*-1 (wherein n = 1 to 17). In still another embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule that is a complement of the nucleotide sequence shown in any of SEQ ID NO:2*n*-1 (wherein n = 1 to 17), or a portion of this nucleotide sequence. A nucleic acid molecule that is complementary to the nucleotide sequence shown in any of SEQ ID NO:2*n*-1 (wherein n = 1 to 17) is one that is sufficiently complementary to the nucleotide sequence shown that it can hydrogen bond with little or no mismatches to the nucleotide sequence shown in of any of SEQ ID NO:2*n*-1 (wherein n = 1 to 17), thereby forming a stable duplex.

As used herein, the term "complementary" refers to Watson-Crick or Hoogsteen base-pairing between nucleotides units of a nucleic acid molecule, whereas the term "binding" is defined as the physical or chemical interaction between two polypeptides or compounds or associated polypeptides or compounds or combinations thereof. Binding includes ionic, non-ionic, Von der Waals, hydrophobic interactions, and the like. A physical interaction can be either direct or indirect. Indirect interactions may be through or due to the effects of another polypeptide or compound. Direct binding refers to interactions that do not take place through, or due to, the effect of another polypeptide or compound, but instead are without other substantial chemical intermediates.

Additionally, the nucleic acid molecule of the invention can comprise only a portion of the nucleic acid sequence of any of SEQ ID NO:2*n*-1 (wherein n = 1 to 17), *e.g.*, a fragment that can be used as a probe or primer, or a fragment encoding a biologically active portion of POLYX. Fragments provided herein are defined as sequences of at least 6 (contiguous) nucleic acids or at least 4 (contiguous) amino acids, a length sufficient to allow for specific hybridization in the case of nucleic acids or for specific recognition of an epitope in the case of amino acids, respectively, and are at most some portion less than a full length sequence. Fragments may be derived from any contiguous portion of a nucleic acid or amino acid sequence of choice. Derivatives are nucleic acid sequences or amino acid sequences formed from the native compounds either directly or by modification or partial substitution. Analogs are nucleic acid sequences or amino acid sequences that have a structure similar to, but not identical to, the native compound but differs from it in respect to certain components or side chains. Analogs may be synthetic or from a different evolutionary origin and may have a similar or opposite metabolic activity compared to wild-type.

Derivatives and analogs may be full-length or other than full-length, if the derivative or analog contains a modified nucleic acid or amino acid, as described *infra*. Derivatives or analogs of the nucleic acids or proteins of the invention include, but are not limited to, molecules comprising regions that are substantially homologous to the nucleic acids or proteins of the invention, in various embodiments, by at least about 70%, 80%, 85%, 90%, 95%, 98%, or even 99% identity (with a preferred identity of 80-99%) over a nucleic acid or amino acid sequence of identical size or when compared to an aligned sequence in which the alignment is done by a computer homology program known in the art, or whose encoding nucleic acid is capable of hybridizing to the complement of a sequence encoding the aforementioned proteins under stringent, moderately stringent, or low stringent conditions. See *e.g.* Ausubel, *et al.*, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY, 1993, and below. An exemplary program is the Gap program (Wisconsin Sequence Analysis Package, Version 8 for UNIX, Genetics Computer Group, University Research Park, Madison, WI) using the default settings, which uses the algorithm of Smith and Waterman (Adv. Appl. Math., 1981, 2: 482-489), which is incorporated herein by reference in its entirety.

As utilized herein, the term "homologous nucleic acid sequence" or "homologous amino acid sequence," or variations thereof, refer to sequences characterized by a homology at the nucleotide level or amino acid level as discussed *supra*. Homologous nucleotide sequences encode those sequences coding for isoforms of POLYX polypeptide. Isoforms can be expressed in different tissues of the same organism as a result of, *e.g.*, alternative splicing of RNA. Alternatively, isoforms can be encoded by different genes. In the invention, homologous nucleotide sequences include nucleotide sequences encoding for a POLYX polypeptide of species other than humans, including, but not limited to, mammals, and thus can include, *e.g.*, mouse, rat, rabbit, dog, cat cow, horse, and other organisms. Homologous nucleotide sequences also include, but are not limited to, naturally occurring allelic variations and mutations of the nucleotide sequences set forth herein. A homologous nucleotide sequence does not, however, include the nucleotide sequence encoding human POLYX protein. Homologous nucleic acid sequences include those nucleic acid sequences that encode conservative amino acid substitutions (see below) in any of SEQ ID NO:2*n* (wherein n = 1 to 17) as well as a polypeptide having POLYX activity. Biological activities of the POLYX proteins are described below. A homologous amino acid sequence does not encode the amino acid sequence of a human POLYX polypeptide.

The nucleotide sequence determined from the cloning of the human POLYX gene allows for the generation of probes and primers designed for use in identifying the cell types disclosed and/or cloning POLYX homologues in other cell types, *e.g.*, from other tissues, as well as POLYX homologues from other mammals. The probe/primer typically comprises a substantially-purified oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, 25, 50, 100, 150, 200, 250, 300, 350 or 400 or more consecutive sense strand nucleotide sequence of SEQ ID NO:2*n*-1 (wherein n = 1 to 17); or an anti-sense strand nucleotide sequence of SEQ ID NO:2*n*-1 (wherein n = 1 to 17); or of a naturally occurring mutant of SEQ ID NO:2*n*-1 (wherein n = 1 to 17).

Probes based upon the human POLYX nucleotide sequence can be used to detect transcripts or genomic sequences encoding the same or homologous proteins. In various embodiments, the probe further comprises a label group attached thereto, *e.g.*, the label group

can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as a part of a diagnostic test kit for identifying cells or tissue which mis-express a POLYX protein, such as by measuring a level of a POLYX-encoding nucleic acid in a sample of cells from a subject *e.g.*, detecting POLYX mRNA levels or determining whether a genomic POLYX gene has been mutated or deleted.

As utilized herein, the term "a polypeptide having a biologically-active portion of POLYX refers to polypeptides exhibiting activity similar, but not necessarily identical to, an activity of a polypeptide of the invention, including mature forms, as measured in a particular biological assay, with or without dose dependency. A nucleic acid fragment encoding a "biologically-active portion of POLYX can be prepared by isolating a portion of SEQ ID NO:2*n*-1 (wherein n = 1 to 17), that encodes a polypeptide having a POLYX biological activity , expressing the encoded portion of POLYX protein (*e.g.*, by recombinant expression *in vitro*), and assessing the activity of the encoded portion of POLY.

POLYX Variants

The invention further encompasses nucleic acid molecules that differ from the disclosed POLYX nucleotide sequences due to degeneracy of the genetic code. These nucleic acids therefore encode the same POLYX protein as those encoded by the nucleotide sequence shown in SEQ ID NO:2*n*-1 (wherein n = 1 to 17). In another embodiment, an isolated nucleic acid molecule of the invention has a nucleotide sequence encoding a protein having an amino acid sequence shown in any of SEQ ID NO:2*n* (wherein n = 1 to 17).

In addition to the human POLYX nucleotide sequence shown in any of SEQ ID NO:2*n*-1 (wherein n = 1 to 17), it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of POLYX may exist within a population (*e.g.*, the human population). Such genetic polymorphism in the POLYX gene may exist among individuals within a population due to natural allelic variation. As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules comprising an open reading frame encoding a POLYX protein, preferably a mammalian POLYX protein. Such natural allelic variations can typically result in 1-5% variance in the nucleotide sequence of the POLYX gene. Any and all such nucleotide variations and resulting

amino acid polymorphisms in POLYX that are the result of natural allelic variation and that do not alter the functional activity of POLYX are intended to be within the scope of the invention.

Additionally, nucleic acid molecules encoding POLYX proteins from other species, and thus that have a nucleotide sequence that differs from the human sequence of any of SEQ ID NO:2*n*-1 (wherein n = 1 to 17), are intended to be within the scope of the invention. Nucleic acid molecules corresponding to natural allelic variants and homologues of the POLYX cDNAs of the invention can be isolated based on their homology to the human POLYX nucleic acids disclosed herein using the human cDNAs, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions.

In another embodiment, an isolated nucleic acid molecule of the invention is at least 6 nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising the nucleotide sequence of any of SEQ ID NO:2*n*-1 (wherein n = 1 to 17). In another embodiment, the nucleic acid is at least 10, 25, 50, 100, 250, 500 or 750 nucleotides in length. In yet another embodiment, an isolated nucleic acid molecule of the invention hybridizes to the coding region. As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least 60% homologous to each other typically remain hybridized to each other.

Homologs (*i.e.*, nucleic acids encoding POLYX proteins derived from species other than human) or other related sequences (*e.g.*, paralogs) can be obtained by low, moderate or high stringency hybridization with all or a portion of the particular human sequence as a probe using methods well known in the art for nucleic acid hybridization and cloning.

As used herein, the phrase "stringent hybridization conditions" refers to conditions under which a probe, primer or oligonucleotide will hybridize to its target sequence, but to no other sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures than shorter sequences. Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The

T_m is the temperature (under defined ionic strength, pH and nucleic acid concentration) at which 50% of the probes complementary to the target sequence hybridize to the target sequence at equilibrium. Since the target sequences are generally present at excess, at T_m, 50% of the probes are occupied at equilibrium. Typically, stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes, primers or oligonucleotides (*e.g.*, 10 nt to 50 nt) and at least about 60°C for longer probes, primers and oligonucleotides. Stringent conditions may also be achieved with the addition of destabilizing agents, such as formamide.

Stringent conditions are known to those skilled in the art and can be found in CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. Preferably, the conditions are such that sequences at least about 65%, 70%, 75%, 85%, 90%, 95%, 98%, or 99% homologous to each other typically remain hybridized to each other. A non-limiting example of stringent hybridization conditions is hybridization in a high salt buffer comprising 6X SSC, 50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.02% BSA, and 500 mg/ml denatured salmon sperm DNA at 65°C. This hybridization is followed by one or more washes in 0.2X SSC, 0.01% BSA at 50°C. An isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to the sequence of any of SEQ ID NO:2*n*-1 (wherein n = 1 to 17) corresponds to a naturally occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (*e.g.*, encodes a natural protein).

In a second embodiment, a nucleic acid sequence that is hybridizable to the nucleic acid molecule comprising the nucleotide sequence of any of SEQ ID NO:2*n*-1 (wherein n = 1 to 17), or fragments, analogs or derivatives thereof, under conditions of moderate stringency is provided. A non-limiting example of moderate stringency hybridization conditions are hybridization in 6X SSC, 5X Denhardt's solution, 0.5% SDS and 100 mg/ml denatured salmon sperm DNA at 55°C, followed by one or more washes in 1X SSC, 0.1% SDS at 37°C. Other conditions of moderate stringency that may be used are well known in the art. See, *e.g.*, Ausubel *et al.* (eds.), 1993, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley &

Sons, NY, and Kriegler, 1990. GENE TRANSFER AND EXPRESSION, A LABORATORY MANUAL, Stockton Press, NY.

In a third embodiment, a nucleic acid that is hybridizable to the nucleic acid molecule comprising the nucleotide sequence of any of SEQ ID NO:2*n*-1 (wherein n = 1 to 17), or fragments, analogs or derivatives thereof, under conditions of low stringency, is provided. A non-limiting example of low stringency hybridization conditions are hybridization in 35% formamide, 5X SSC, 50 mM Tris-HCl (pH 7.5), 5 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.2% BSA, 100 mg/ml denatured salmon sperm DNA, 10% (wt/vol) dextran sulfate at 40°C, followed by one or more washes in 2X SSC, 25 mM Tris-HCl (pH 7.4), 5 mM EDTA, and 0.1% SDS at 50°C. Other conditions of low stringency that may be used are well known in the art (*e.g.*, as employed for cross-species hybridizations). *See, e.g.*, Ausubel, *et al.*, (eds.), 1993. CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, NY, and Kriegler, 1990. GENE TRANSFER AND EXPRESSION, A LABORATORY MANUAL, Stockton Press, NY; Shilo and Weinberg, 1981. *Proc. Natl. Acad. Sci. USA 78*: 6789-6792.

Conservative Mutations

In addition to naturally-occurring allelic variants of the POLYX sequence that may exist in the population, the skilled artisan will further appreciate that changes can be introduced by mutation into the nucleotide sequence of any of SEQ ID NO:2*n*-1 (wherein n = 1 to 17), thereby leading to changes in the amino acid sequence of the encoded POLYX protein, without altering the functional ability of the POLYX protein. For example, nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues can be made in the sequence of any of SEQ ID NO:2n-1 (wherein n = 1 to 17). A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequence of POLYX without altering the biological activity, whereas an "essential" amino acid residue is required for biological activity. For example, amino acid residues that are conserved among the POLYX proteins of the invention, are predicted to be particularly non-amenable to such alteration.

Amino acid residues that are conserved among members of a POLYX family are predicted to be less amenable to alteration. For example, a POLYX protein according to the

invention can contain at least one domain that is a typically conserved region in a POLYX family member. As such, these conserved domains are not likely to be amenable to mutation. Other amino acid residues, however, (*e.g.*, those that are not conserved or only semi-conserved among members of the POLYX family) may not be as essential for activity and thus are more likely to be amenable to alteration.

Another aspect of the invention pertains to nucleic acid molecules encoding POLYX proteins that contain changes in amino acid residues that are not essential for activity. Such POLYX proteins differ in amino acid sequence from any of any of SEQ ID NO:2*n* (wherein n = 1 to 17), yet retain biological activity. In one embodiment, the isolated nucleic acid molecule comprises a nucleotide sequence encoding a protein, wherein the protein comprises an amino acid sequence at least about 75% homologous to the amino acid sequence of any of SEQ ID NO:2*n* (wherein n = 1 to 17). Preferably, the protein encoded by the nucleic acid is at least about 80% homologous to any of SEQ ID NO:2*n* (wherein n = 1 to 17), more preferably at least about 90%, 95%, 98%, and most preferably at least about 99% homologous to SEQ ID NO:2*n* (wherein n = 1 to 17).

An isolated nucleic acid molecule encoding a POLYX protein homologous to the protein of any of SEQ ID NO:2*n* (wherein n = 1 to 17) can be created by introducing one or more nucleotide substitutions, additions or deletions into the corresponding nucleotide sequence (*i.e.,* SEQ ID NO:2*n*-1 for the corresponding *n*), such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein.

Mutations can be introduced into SEQ ID NO:2*n*-1 (wherein n = 1 to 17) by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (*e.g.*, lysine, arginine, histidine), acidic side chains (*e.g.*, aspartic acid, glutamic acid), uncharged polar side chains (*e.g.*, glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (*e.g.*, alanine, valine, leucine, isoleucine,

proline, phenylalanine, methionine, tryptophan), β-branched side chains (*e.g.*, threonine, valine, isoleucine) and aromatic side chains (*e.g.*, tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted nonessential amino acid residue in POLYX is replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of a POLYX coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for POLYX biological activity to identify mutants that retain activity. Following mutagenesis of SEQ ID NO:2*n*-1 (wherein n = 1 to 17), the encoded protein can be expressed by any recombinant technology known in the art and the activity of the protein can be determined.

In one embodiment, a mutant POLYX protein can be assayed for: (*i*) the ability to form protein:protein interactions with other POLYX proteins, other cell-surface proteins, or biologically-active portions thereof; (*ii*) complex formation between a mutant POLYX protein and a POLYX receptor; (*iii*) the ability of a mutant POLYX protein to bind to an intracellular target protein or biologically active portion thereof; (*e.g* , avidin proteins); (*iv*) the ability to bind BRA protein; or (*v*) the ability to specifically bind an anti-POLYX protein antibody.

Antisense Nucleic Acids

Another aspect of the invention pertains to isolated antisense nucleic acid molecules that are hybridizable to or complementary to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:2*n*-1 (wherein n = 1 to 17), or fragments, analogs or derivatives thereof. An "antisense" nucleic acid comprises a nucleotide sequence that is complementary to a "sense" nucleic acid encoding a protein, *e.g.*, complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. In specific aspects, antisense nucleic acid molecules are provided that comprise a sequence complementary to at least about 10, 25, 50, 100, 250 or 500 nucleotides or an entire POLYX coding strand, or to only a portion thereof. Nucleic acid molecules encoding fragments, homologs, derivatives and analogs of a POLYX protein of any of SEQ ID NO:2*n* (wherein n = 1 to 17) or antisense nucleic acids complementary to a POLYX nucleic acid sequence of SEQ ID NO:2n-1 (wherein n = 1 to 17) are additionally provided.

In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence encoding POLY. The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues (*e.g.*, the protein coding region of a human POLYX that corresponds to any of SEQ ID NO:2*n* (wherein n = 1 to 17)). In another embodiment, the antisense nucleic acid molecule is antisense to a "non-coding region" of the coding strand of a nucleotide sequence encoding POLY. The term "non-coding region" refers to 5' and 3' sequences which flank the coding region that are not translated into amino acids (*i.e.*, also referred to as 5' and 3' non-translated regions).

Given the coding strand sequences encoding POLYX disclosed herein (*e.g.*, SEQ ID NO:2*n*-1 (wherein n = 1 to 17)), antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick or Hoogsteen base-pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of POLYX mRNA, but more preferably is an oligonucleotide that is antisense to only a portion of the coding or non-coding region of POLYX mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of POLYX mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis or enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (*e.g.*, an antisense oligonucleotide) can be chemically synthesized using naturally-occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, *e.g.*, phosphorothioate derivatives and acridine-substituted nucleotides can be used.

Examples of modified nucleotides that can be used to generate the antisense nucleic acid include: 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxylmethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine,

7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (*i.e.*, RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a POLYX protein to thereby inhibit expression of the protein, *e.g.*, by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule that binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of antisense nucleic acid molecules of the invention includes direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface (*e.g.*, by linking the antisense nucleic acid molecules to peptides or antibodies that bind to cell surface receptors or antigens). The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

In yet another embodiment, the antisense nucleic acid molecule of the invention is an α-anomeric nucleic acid molecule. An α-anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual α-units, the strands run parallel to each other (Gaultier, *et al.*, 1987. *Nucl. Acids Res.* 15: 6625-6641). The

antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (Inoue, *et al.*, 1987. *Nucl. Acids Res.* 15: 6131-6148) or a chimeric RNA-DNA analogue (Inoue, *et al.*, 1987. *FEBS Lett.* 215: 327-330).

Ribozymes and PNA Moieties

Such modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject.

In still another embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity that are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (*e.g.*, hammerhead ribozymes; described by Haselhoff and Gerlach, 1988. *Nature* 334: 585-591) can be used to catalytically-cleave POLYX mRNA transcripts to thereby inhibit translation of POLYX mRNA. A ribozyme having specificity for a POLYX-encoding nucleic acid can be designed based upon the nucleotide sequence of a POLYX DNA disclosed herein (*i.e.*, SEQ ID NO:2*n*-1 (wherein n = 1 to 17)). For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a POLYX-encoding mRNA. See, *e.g.*, Cech, *et al.*, U.S. Patent No. 4,987,071; and Cech, *et al.*, U.S. Patent No. 5,116,742. Alternatively, POLYX mRNA can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules (Bartel, *et al.*, 1993. *Science* 261: 1411-1418).

Alternatively, POLYX gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the POLYX (*e.g.*, the POLYX promoter and/or enhancers) to form triple helical structures that prevent transcription of the POLYX gene in target cells. *See, e.g.,* Helene, 1991. *Anticancer Drug Des.* 6: 569-84; Helene, *et al.*, 1992. *Ann. N.Y. Acad. Sci.* 660: 27-36; and Maher, 1992. *Bioassays* 14: 807-15.

In various embodiments, the nucleic acids of POLYX can be modified at the base moiety, sugar moiety or phosphate backbone to improve, *e.g.*, the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids (Hyrup, *et al.*, 1996. *Bioorg. Med. Chem.* 4: 5-23). As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics, *e.g.*, DNA mimics, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup, *et al.*, 1996. *supra*; Perry-O'Keefe, *et al.*, 1996. *Proc. Natl. Acad. Sci. USA* 93: 14670-14675.

PNAs of POLYX can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, *e.g.*, inducing transcription or translation arrest or inhibiting replication. PNAs of POLYX can also be used, *e.g.*, in the analysis of single base pair mutations in a gene by, *e.g.*, PNA directed PCR clamping; as artificial restriction enzymes when used in combination with other enzymes, *e.g.*, S1 nucleases (*see,* Hyrup, 1996., *supra*); or as probes or primers for DNA sequence and hybridization (*see,* Hyrup, *et al.*, 1996.; Perry-O'Keefe, 1996., *supra*).

In another embodiment, PNAs of POLYX can be modified, *e.g.*, to enhance their stability or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras of POLYX can be generated that may combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes, *e.g.*, RNase H and DNA polymerases, to interact with the DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleobases, and orientation (*see,* Hyrup, 1996., *supra*). The synthesis of PNA-DNA chimeras can be performed as described in Finn, *et al.*, (1996. *Nucl. Acids Res.* 24: 3357-3363). For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry, and modified nucleoside analogs, *e.g.*,

5'-(4-methoxytrityl)amino-5'-deoxy-thymidine phosphoramidite, can be used between the PNA and the 5' end of DNA (Mag, *et al.*, 1989. *Nucl. Acid Res.* 17: 5973-5988). PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment (*see*, Finn, *et al.*, 1996., *supra*). Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment. *See, e.g.*, Petersen, *et al.*, 1975. *Bioorg. Med. Chem. Lett.* 5: 1119-11124.

In other embodiments, the oligonucleotide may include other appended groups such as peptides (*e.g.*, for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (*see, e.g.*, Letsinger, *et al.*, 1989. *Proc. Natl. Acad. Sci. U.S.A.* 86: 6553-6556; Lemaitre, *et al.*, 1987. *Proc. Natl. Acad. Sci.* 84: 648-652; PCT Publication No. WO88/09810) or the blood-brain barrier (*see, e.g.*, PCT Publication No. WO 89/10134). In addition, oligonucleotides can be modified with hybridization triggered cleavage agents (*see, e.g.*, Krol, *et al.*, 1988. *BioTechniques* 6:958-976) or intercalating agents (*see, e.g.*, Zon, 1988. *Pharm. Res.* 5: 539-549). To this end, the oligonucleotide may be conjugated to another molecule, *e.g.*, a peptide, a hybridization triggered cross-linking agent, a transport agent, a hybridization-triggered cleavage agent, and the like.

Characterization of POLYX Polypeptides

A polypeptide according to the invention includes a polypeptide including the amino acid sequence of POLYX polypeptides whose sequences are provided in any SEQ ID NO:*2n* (wherein n = 1 to 17) and includes SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32 and/or 34. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residues shown in SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32 and/or 34, while still encoding a protein that maintains its POLYX activities and physiological functions, or a functional fragment thereof.

In general, a POLYX variant that preserves POLYX-like function includes any variant in which residues at a particular position in the sequence have been substituted by other amino acids, and further include the possibility of inserting an additional residue or residues between two residues of the parent protein as well as the possibility of deleting one or more residues from the parent sequence. Any amino acid substitution, insertion, or deletion is encompassed

by the invention. In favorable circumstances, the substitution is a conservative substitution as defined above.

One aspect of the invention pertains to isolated POLYX proteins, and biologically-active portions thereof, or derivatives, fragments, analogs or homologs thereof. Also provided are polypeptide fragments suitable for use as immunogens to raise anti-POLYX antibodies. In one embodiment, native POLYX proteins can be isolated from cells or tissue sources by an appropriate purification scheme using standard protein purification techniques. In another embodiment, POLYX proteins are produced by recombinant DNA techniques. Alternative to recombinant expression, a POLYX protein or polypeptide can be synthesized chemically using standard peptide synthesis techniques.

An "isolated" or "purified" polypeptide or protein or biologically-active portion thereof is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the POLYX protein is derived, or substantially free from chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of POLYX proteins in which the protein is separated from cellular components of the cells from which it is isolated or recombinantly-produced. In one embodiment, the language "substantially free of cellular material" includes preparations of POLYX proteins having less than about 30% (by dry weight) of non-POLYX proteins (also referred to herein as a "contaminating protein"), more preferably less than about 20% of non-POLYX proteins, still more preferably less than about 10% of non-POLYX proteins, and most preferably less than about 5% of non-POLYX proteins. When the POLYX protein or biologically-active portion thereof is recombinantly-produced, it is also preferably substantially free of culture medium, *i.e.*, culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the POLYX protein preparation.

As utilized herein, the phrase "substantially free of chemical precursors or other chemicals" includes preparations of POLYX protein in which the protein is separated from chemical precursors or other chemicals that are involved in the synthesis of the protein. In one embodiment, the language "substantially free of chemical precursors or other chemicals"

includes preparations of POLYX protein having less than about 30% (by dry weight) of chemical precursors or non-POLYX chemicals, more preferably less than about 20% chemical precursors or non-POLYX chemicals, still more preferably less than about 10% chemical precursors or non-POLYX chemicals, and most preferably less than about 5% chemical precursors or non-POLYX chemicals.

Biologically-active portions of a POLYX protein include peptides comprising amino acid sequences sufficiently homologous to or derived from the amino acid sequence of the POLYX protein which include fewer amino acids than the full-length POLYX proteins, and exhibit at least one activity of a POLYX protein. Typically, biologically-active portions comprise a domain or motif with at least one activity of the POLYX protein. A biologically-active portion of a POLYX protein can be a polypeptide which is, for example, 10, 25, 50, 100 or more amino acids in length.

A biologically-active portion of a POLYX protein of the invention may contain at least one of the above-identified conserved domains. Moreover, other biologically active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the functional activities of a native POLYX protein.

In an embodiment, the POLYX protein has an amino acid sequence shown in any of SEQ ID NO:2*n* (wherein n = 1 to 17). In other embodiments, the POLYX protein is substantially homologous to any of SEQ ID NO:2*n* (wherein n = 1 to 17) and retains the functional activity of the protein of any of SEQ ID NO:2*n* (wherein n = 1 to 17), yet differs in amino acid sequence due to natural allelic variation or mutagenesis, as described in detail below. Accordingly, in another embodiment, the POLYX protein is a protein that comprises an amino acid sequence at least about 45% homologous, and more preferably about 55, 65, 70, 75, 80, 85, 90, 95, 98 or even 99% homologous to the amino acid sequence of any of SEQ ID NO:2*n* (wherein n = 1 to 17) and retains the functional activity of the POLYX proteins of the corresponding polypeptide having the sequence of SEQ ID NO:2*n* (wherein n = 1 to 17).

Determining Homology Between Two or More Sequences

To determine the percent homology of two amino acid sequences or of two nucleic acids, the sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in the sequence of a first amino acid or nucleic acid sequence for optimal alignment with a second amino or nucleic acid sequence). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are homologous at that position (*i.e.*, as used herein amino acid or nucleic acid "homology" is equivalent to amino acid or nucleic acid "identity").

The nucleic acid sequence homology may be determined as the degree of identity between two sequences. The homology may be determined using computer programs known in the art, such as GAP software provided in the GCG program package. *See,* Needleman and Wunsch, 1970. *J. Mol. Biol.* 48: 443-453. Using GCG GAP software with the following settings for nucleic acid sequence comparison: GAP creation penalty of 5.0 and GAP extension penalty of 0.3, the coding region of the analogous nucleic acid sequences referred to above exhibits a degree of identity preferably of at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99%, with the CDS (encoding) part of the DNA sequence shown in SEQ ID NO:2*n*-1 (wherein n = 1 to 17), *e.g.*, SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31 and/or 33.

The term "sequence identity" refers to the degree to which two polynucleotide or polypeptide sequences are identical on a residue-by-residue basis over a particular region of comparison. The term "percentage of sequence identity" is calculated by comparing two optimally aligned sequences over that region of comparison, determining the number of positions at which the identical nucleic acid base (*e.g.*, A, T, C, G, U, or I, in the case of nucleic acids) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the region of comparison (*i.e.*, the window size), and multiplying the result by 100 to yield the percentage of sequence identity. The term "substantial identity" as used herein denotes a characteristic of a

polynucleotide sequence, wherein the polynucleotide comprises a sequence that has at least 80 percent sequence identity, preferably at least 85 percent identity and often 90 to 95 percent sequence identity, more usually at least 99 percent sequence identity as compared to a reference sequence over a comparison region.

Chimeric and Fusion Proteins

The invention also provides POLYX chimeric or fusion proteins. As used herein, a POLYX "chimeric protein" or "fusion protein" comprises a POLYX polypeptide operatively-linked to a non-POLYX polypeptide. An "POLYX polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a POLYX protein shown in SEQ ID NO:2*n* (wherein n = 1 to 17), [*e.g.*, SEQ ID NO:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32 and/or 34], whereas a "non-POLYX polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a protein that is not substantially homologous to the POLYX protein (*e.g.*, a protein that is different from the POLYX protein and that is derived from the same or a different organism). Within a POLYX fusion protein the POLYX polypeptide can correspond to all or a portion of a POLYX protein. In one embodiment, a POLYX fusion protein comprises at least one biologically-active portion of a POLYX protein. In another embodiment, a POLYX fusion protein comprises at least two biologically-active portions of a POLYX protein. In yet another embodiment, a POLYX fusion protein comprises at least three biologically-active portions of a POLYX protein. Within the fusion protein, the term "operatively-linked" is intended to indicate that the POLYX polypeptide and the non-POLYX polypeptide are fused in-frame with one another. The non-POLYX polypeptide can be fused to the amino-terminus or carboxyl-terminus of the POLYX polypeptide.

In one embodiment, the fusion protein is a GST-POLYX fusion protein in which the POLYX sequences are fused to the carboxyl-terminus of the GST (glutathione S-transferase) sequences. Such fusion proteins can facilitate the purification of recombinant POLYX polypeptides.

In another embodiment, the fusion protein is a POLYX protein containing a heterologous signal sequence at its amino-terminus. In certain host cells (*e.g.*, mammalian

host cells), expression and/or secretion of POLYX can be increased through use of a heterologous signal sequence.

In yet another embodiment, the fusion protein is a POLYX-immunoglobulin fusion protein in which the POLYX sequences are fused to sequences derived from a member of the immunoglobulin protein family. The POLYX-immunoglobulin fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject to inhibit an interaction between a POLYX ligand and a POLYX protein on the surface of a cell, to thereby suppress POLYX-mediated signal transduction *in vivo*. The POLYX-immunoglobulin fusion proteins can be used to affect the bioavailability of a POLYX cognate ligand. Inhibition of the POLYX ligand/POLYX interaction may be useful therapeutically for both the treatment of proliferative and differentiative disorders, as well as modulating (*e.g.*, promoting or inhibiting) cell survival. Moreover, the POLYX-immunoglobulin fusion proteins of the invention can be used as immunogens to produce anti-POLYX antibodies in a subject, to purify POLYX ligands, and in screening assays to identify molecules that inhibit the interaction of POLYX with a POLYX ligand.

A POLYX chimeric or fusion protein of the invention can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, *e.g.*, by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers that give rise to complementary overhangs between two consecutive gene fragments that can subsequently be annealed and reamplified to generate a chimeric gene sequence (*see, e.g.,* Ausubel, *et al.* (eds.) CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (*e.g.*, a GST polypeptide). A POLYX-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the POLYX protein.

POLYX Agonists and Antagonists

The invention also pertains to variants of the POLYX proteins that function as either POLYX agonists (*i.e.*, mimetics) or as POLYX antagonists. Variants of the POLYX protein can be generated by mutagenesis (*e.g.*, discrete point mutation or truncation of the POLYX protein). An agonist of a POLYX protein can retain substantially the same, or a subset of, the biological activities of the naturally-occurring form of a POLYX protein. An antagonist of a POLYX protein can inhibit one or more of the activities of the naturally occurring form of a POLYX protein by, for example, competitively binding to a downstream or upstream member of a cellular signaling cascade which includes the POLYX protein. Thus, specific biological effects can be elicited by treatment with a variant of limited function. In one embodiment, treatment of a subject with a variant having a subset of the biological activities of the naturally occurring form of the protein has fewer side effects in a subject relative to treatment with the naturally occurring form of the POLYX proteins.

Variants of the POLYX proteins that function as either POLYX agonists (*i.e.*, mimetics) or as POLYX antagonists can be identified by screening combinatorial libraries of mutants (*e.g.*, truncation mutants) of the POLYX proteins for POLYX protein agonist or antagonist activity. In one embodiment, a variegated library of POLYX variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of POLYX variants can be produced by, for example, enzymatically-ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential POLYX sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (*e.g.*, for phage display) containing the set of POLYX sequences therein. There are a variety of methods which can be used to produce libraries of potential POLYX variants from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be performed in an automatic DNA synthesizer, and the synthetic gene then ligated into an appropriate expression vector. Use of a degenerate set of genes allows for the provision, in one mixture, of all of the sequences encoding the desired set of potential POLYX sequences. Methods for synthesizing degenerate oligonucleotides are well-known within the art. *See, e.g.,* Narang, 1983. *Tetrahedron* 39: 3;

Itakura, *et al.*, 1984. *Annu. Rev. Biochem.* 53: 323; Itakura, *et al.*, 1984. *Science* 198: 1056; Ike, *et al.*, 1983. *Nucl. Acids Res.* 11: 477.

Polypeptide Libraries

In addition, libraries of fragments of the POLYX protein coding sequences can be used to generate a variegated population of POLYX fragments for screening and subsequent selection of variants of a POLYX protein. In one embodiment, a library of coding sequence fragments can be generated by treating a double-stranded PCR fragment of a POLYX coding sequence with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double-stranded DNA that can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S_1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, expression libraries can be derived which encodes amino-terminal and internal fragments of various sizes of the POLYX proteins.

Various techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. Such techniques are adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of POLYX proteins. The most widely used techniques, which are amenable to high throughput analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recursive ensemble mutagenesis (REM), a new technique that enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify POLYX variants. *See, e.g.,* Arkin and Yourvan, 1992. *Proc. Natl. Acad. Sci. USA* 89: 7811-7815; Delgrave, *et al.*, 1993. *Protein Engineering* 6:327-331.

Anti-POLYX Antibodies

The invention encompasses antibodies and antibody fragments, such as F_{ab} or $(F_{ab})_2$, that bind immunospecifically to any of the POLYX polypeptides of said invention.

An isolated POLYX protein, or a portion or fragment thereof, can be used as an immunogen to generate antibodies that bind to POLYX polypeptides using standard techniques for polyclonal and monoclonal antibody preparation. The full-length POLYX proteins can be used or, alternatively, the invention provides antigenic peptide fragments of POLYX proteins for use as immunogens. The antigenic POLYX peptides comprises at least 4 amino acid residues of the amino acid sequence shown in SEQ ID NO:2*n* (wherein n = 1 to 17) and encompasses an epitope of POLYX such that an antibody raised against the peptide forms a specific immune complex with POLY. Preferably, the antigenic peptide comprises at least 6, 8, 10, 15, 20, or 30 amino acid residues. Longer antigenic peptides are sometimes preferable over shorter antigenic peptides, depending on use and according to methods well known to someone skilled in the art.

In certain embodiments of the invention, at least one epitope encompassed by the antigenic peptide is a region of POLYX that is located on the surface of the protein (*e.g.*, a hydrophilic region). As a means for targeting antibody production, hydropathy plots showing regions of hydrophilicity and hydrophobicity may be generated by any method well known in the art, including, for example, the Kyte-Doolittle or the Hopp-Woods methods, either with or without Fourier transformation (*see, e.g.*, Hopp and Woods, 1981. *Proc. Nat. Acad. Sci. USA* 78: 3824-3828; Kyte and Doolittle, 1982. *J. Mol. Biol.* 157: 105-142, each incorporated herein by reference in their entirety).

As disclosed herein, POLYX protein sequences of SEQ ID NO:2*n* (wherein n = 1 to 17), or derivatives, fragments, analogs, or homologs thereof, may be utilized as immunogens in the generation of antibodies that immunospecifically-bind these protein components. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically-active portions of immunoglobulin molecules, *i.e.*, molecules that contain an antigen binding site that specifically-binds (immunoreacts with) an antigen, such as POLY. Such antibodies include, but are not limited to, polyclonal, monoclonal, chimeric, single chain, F_{ab} and $F_{(ab')2}$

fragments, and an F_{ab} expression library. In a specific embodiment, antibodies to human POLYX proteins are disclosed. Various procedures known within the art may be used for the production of polyclonal or monoclonal antibodies to a POLYX protein sequence of SEQ ID NO:2*n* (wherein n = 1 to 17), or a derivative, fragment, analog, or homolog thereof. Some of these proteins are discussed, *infra*.

For the production of polyclonal antibodies, various suitable host animals (*e.g.*, rabbit, goat, mouse or other mammal) may be immunized by injection with the native protein, or a synthetic variant thereof, or a derivative of the foregoing. An appropriate immunogenic preparation can contain, for example, recombinantly-expressed POLYX protein or a chemically-synthesized POLYX polypeptide. The preparation can further include an adjuvant. Various adjuvants used to increase the immunological response include, but are not limited to, Freund's (complete and incomplete), mineral gels (*e.g.*, aluminum hydroxide), surface active substances (*e.g.*, lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, dinitrophenol, etc.), human adjuvants such as *Bacille Calmette-Guerin* and *Corynebacterium parvum*, or similar immunostimulatory agents. If desired, the antibody molecules directed against POLYX can be isolated from the mammal (*e.g.*, from the blood) and further purified by well known techniques, such as protein A chromatography to obtain the IgG fraction.

The term "monoclonal antibody" or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one species of an antigen binding site capable of immunoreacting with a particular epitope of POLY. A monoclonal antibody composition thus typically displays a single binding affinity for a particular POLYX protein with which it immunoreacts. For preparation of monoclonal antibodies directed towards a particular POLYX protein, or derivatives, fragments, analogs or homologs thereof, any technique that provides for the production of antibody molecules by continuous cell line culture may be utilized. Such techniques include, but are not limited to, the hybridoma technique (*see, e.g.,* Kohler & Milstein, 1975. *Nature* 256: 495-497); the trioma technique; the human B-cell hybridoma technique (*see, e.g.,* Kozbor, *et al.,* 1983. *Immunol. Today* 4: 72) and the EBV hybridoma technique to produce human monoclonal antibodies (*see, e.g.,* Cole, *et al.,* 1985. In: MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96). Human monoclonal antibodies may be utilized in the practice of the invention and may be

produced by using human hybridomas (*see, e.g.*. Cote, *et al.*, 1983. *Proc. Natl. Acad. Sci. USA* 80: 2026-2030) or by transforming human B-cells with Epstein Barr Virus *in vitro* (*see, e.g.*, Cole, *et al.*, 1985. In: MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96). Each of the above citations is incorporated herein by reference in their entirety.

According to the invention, techniques can be adapted for the production of single-chain antibodies specific to a POLYX protein (*see, e.g.*, U.S. Patent No. 4,946,778). In addition, methods can be adapted for the construction of F_{ab} expression libraries (*see, e.g.*, Huse, *et al.*, 1989. *Science* 246: 1275-1281) to allow rapid and effective identification of monoclonal F_{ab} fragments with the desired specificity for a POLYX protein or derivatives, fragments, analogs or homologs thereof. Non-human antibodies can be "humanized" by techniques well known in the art. *See, e.g.*, U.S. Patent No. 5,225,539. Antibody fragments that contain the idiotypes to a POLYX protein may be produced by techniques known in the art including, but not limited to: (*i*) an $F_{(ab')2}$ fragment produced by pepsin digestion of an antibody molecule; (*ii*) an F_{ab} fragment generated by reducing the disulfide bridges of an $F_{(ab')2}$ fragment; (*iii*) an F_{ab} fragment generated by the treatment of the antibody molecule with papain and a reducing agent and (*iv*) F_v fragments.

Additionally, recombinant anti-POLYX antibodies, such as chimeric and humanized monoclonal antibodies, comprising both human and non-human portions, which can be made using standard recombinant DNA techniques, are within the scope of the invention. Such chimeric and humanized monoclonal antibodies can be produced by recombinant DNA techniques known in the art, for example using methods described in International Application No. PCT/US86/02269; European Patent Application No. 184,187; European Patent Application No. 171,496; European Patent Application No. 173,494; PCT International Publication No. WO 86/01533; U.S. Patent No. 4,816,567; U.S. Pat. No. 5,225,539; European Patent Application No. 125,023; Better, *et al.*, 1988. *Science* 240: 1041-1043; Liu, *et al.*, 1987. *Proc. Natl. Acad. Sci. USA* 84: 3439-3443; Liu, *et al.*, 1987. *J. Immunol.* 139: 3521-3526; Sun, *et al.*, 1987. *Proc. Natl. Acad. Sci. USA* 84: 214-218; Nishimura, *et al.*, 1987. *Cancer Res.* 47: 999-1005; Wood, *et al.*, 1985. *Nature* 314 :446-449; Shaw, *et al.*, 1988. *J. Natl. Cancer Inst.* 80: 1553-1559); Morrison(1985) *Science* 229:1202-1207; Oi, *et al.* (1986) *BioTechniques* 4:214; Jones, *et al.*, 1986. *Nature* 321: 552-525; Verhoeyan, *et al.*, 1988. *Science* 239: 1534;

and Beidler, *et al.*, 1988. *J. Immunol.* 141: 4053-4060. Each of the above citations are incorporated herein by reference in their entirety.

In one embodiment, methods for the screening of antibodies that possess the desired specificity include, but are not limited to, enzyme-linked immunosorbent assay (ELISA) and other immunologically-mediated techniques known within the art. In a specific embodiment, selection of antibodies that are specific to a particular domain of a POLYX protein is facilitated by generation of hybridomas that bind to the fragment of a POLYX protein possessing such a domain. Thus, antibodies that are specific for a desired domain within a POLYX protein, or derivatives, fragments, analogs or homologs thereof, are also provided herein.

Anti-POLYX antibodies may be used in methods known within the art relating to the localization and/or quantitation of a POLYX protein (*e.g.,* for use in measuring levels of the POLYX protein within appropriate physiological samples, for use in diagnostic methods, for use in imaging the protein, and the like). In a given embodiment, antibodies for POLYX proteins, or derivatives, fragments, analogs or homologs thereof, that contain the antibody derived binding domain, are utilized as pharmacologically-active compounds (hereinafter "Therapeutics").

An anti-POLYX antibody (*e.g.*, monoclonal antibody) can be used to isolate a POLYX polypeptide by standard techniques, such as affinity chromatography or immunoprecipitation. An anti-POLYX antibody can facilitate the purification of natural POLYX polypeptide from cells and of recombinantly-produced POLYX polypeptide expressed in host cells. Moreover, an anti-POLYX antibody can be used to detect POLYX protein (*e.g.*, in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the POLYX protein. Anti-POLYX antibodies can be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, *e.g.*, to, for example, determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling (*i.e.*, physically linking) the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish

peroxidase, alkaline phosphatase, β-galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I, ^{131}I, ^{35}S or ^{3}H.

POLYX Recombinant Expression Vectors and Host Cells

Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding a POLYX protein, or derivatives, fragments, analogs or homologs thereof. As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (*e.g.*, non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively-linked. Such vectors are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present Specification, "plasmid" and "vector" can be used interchangeably, as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (*e.g.*, replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected on the

basis of the host cells to be used for expression, that is operatively-linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably-linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner that allows for expression of the nucleotide sequence (*e.g.*, in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell).

As utilized herein, the phrase "regulatory sequence" is intended to includes promoters, enhancers and other expression control elements (*e.g.*, polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, CA (1990). Regulatory sequences include those that direct constitutive expression of a nucleotide sequence in many types of host cell and those that direct expression of the nucleotide sequence only in certain host cells (*e.g.*, tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, etc. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (*e.g.*, POLYX proteins, mutant forms of POLYX proteins, fusion proteins, etc.).

The recombinant expression vectors of the invention can be designed for expression of POLYX proteins in prokaryotic or eukaryotic cells. For example, POLYX proteins can be expressed in bacterial cells such as *Escherichia coli*, insect cells (using baculovirus expression vectors) yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T_7 promoter regulatory sequences and T_7 polymerase.

Expression of proteins in prokaryotes is most often carried out in *Escherichia coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded

therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: (*i*) to increase expression of recombinant protein; (*ii*) to increase the solubility of the recombinant protein; and (*iii*) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor X_a, thrombin, and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith and Johnson, 1988. *Gene* 67: 31-40), pMAL (New England Biolabs, Beverly, Mass.) and pRIT5 (Pharmacia, Piscataway, NJ) that fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

Examples of suitable inducible non-fusion *Escherichia coli* expression vectors include pTrc (Amrann *et al.*, (1988) *Gene* 69:301-315) and pET 11d (Studier, *et al.*, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 60-89).

One strategy to maximize recombinant protein expression in *Escherichia coli* is to express the protein in a host bacteria with an impaired capacity to proteolytically-cleave the recombinant protein. *See, e.g.*. Gottesman, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 119-128. Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in *Escherichia coli* (*see, e.g.,* Wada, *et al.*, 1992. *Nucl. Acids Res.* 20: 2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

In another embodiment, the POLYX expression vector is a yeast expression vector. Examples of vectors for expression in yeast *Saccharomyces cerivisae* include pYepSec1 (Baldari, *et al.*, 1987. *EMBO J.* 6: 229-234), pMFa (Kurjan and Herskowitz, 1982. *Cell* 30: 933-943), pJRY88 (Schultz *et al.*, 1987. *Gene* 54: 113-123), pYES2 (Invitrogen Corporation, San Diego, Calif.), and picZ (InVitrogen Corp, San Diego, Calif.).

Alternatively, POLYX can be expressed in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (*e.g.*, SF9 cells) include the pAc series (Smith, *et al.*, 1983. *Mol. Cell. Biol.* 3: 2156-2165) and the pVL series (Lucklow and Summers, 1989. *Virology* 170: 31-39).

In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, 1987. *Nature* 329: 840) and pMT2PC (Kaufman, *et al.*, 1987. *EMBO J.* 6: 187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, adenovirus 2, cytomegalovirus, and simian virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells *see, e.g.*, Chapters 16 and 17 of Sambrook, *et al.*, MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989.

In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (*e.g.*, tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; *see,* Pinkert, *et al.*, 1987. *Genes Dev.* 1: 268-277), lymphoid-specific promoters (*see,* Calame and Eaton, 1988. *Adv. Immunol.* 43: 235-275), in particular promoters of T cell receptors (*see,* Winoto and Baltimore, 1989. *EMBO J.* 8: 729-733) and immunoglobulins (*see,* Banerji, *et al.*, 1983. *Cell* 33: 729-740; Queen and Baltimore, 1983. *Cell* 33: 741-748), neuron-specific promoters (*e.g.*, the neurofilament promoter; *see,* Byrne and Ruddle, 1989. *Proc. Natl. Acad. Sci. USA* 86: 5473-5477), pancreas-specific promoters (*see,* Edlund, *et al.*, 1985. *Science* 230: 912-916), and mammary gland-specific promoters (*e.g.*, milk whey promoter; U.S. Pat. No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, *e.g.*, the murine hox promoters (Kessel and Gruss, 1990. *Science* 249: 374-379) and the α-fetoprotein promoter (*see,* Campes and Tilghman, 1989. *Genes Dev.* 3: 537-546).

The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operatively-linked to a regulatory sequence in a manner that allows for expression (by transcription of the DNA molecule) of an RNA molecule that is antisense to POLYX mRNA. Regulatory sequences operatively linked to a nucleic acid cloned in the antisense orientation can be chosen that direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen that direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes *see, e.g.,* Weintraub, *et al.*, "Antisense RNA as a molecular tool for genetic analysis," *Reviews-Trends in Genetics*, Vol. 1(1) 1986.

Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but also to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic or eukaryotic cell. For example, POLYX protein can be expressed in bacterial cells such as *Escherichia coli*, insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (*e.g.*, DNA) into a host cell, including calcium phosphate or calcium

chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, *et al.* (MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989), and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (*e.g.*, resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Various selectable markers include those that confer resistance to drugs, such as G418, hygromycin, and methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding POLYX or can be introduced on a separate vector. Cells stably-transfected with the introduced nucleic acid can be identified by drug selection (*e.g.*, cells that have incorporated the selectable marker gene will survive, while the other cells die).

A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (*i.e.*, express) POLYX protein. Accordingly, the invention further provides methods for producing POLYX protein using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of invention (*i.e.*, into which a recombinant expression vector encoding POLYX protein has been introduced) in a suitable medium such that POLYX protein is produced. In another embodiment, the method further comprises isolating POLYX protein from the medium or the host cell.

Transgenic Animals

The host cells of the invention can also be used to produce non-human transgenic animals. For example, in one embodiment, a host cell of the invention is a fertilized oocyte or an embryonic stem cell into which POLYX protein-coding sequences have been introduced. These host cells can then be used to create non-human transgenic animals in which exogenous POLYX sequences have been introduced into their genome or homologous recombinant animals in which endogenous POLYX sequences have been altered. Such animals are useful

for studying the function and/or activity of POLYX protein and for identifying and/or evaluating modulators of POLYX protein activity. As used herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, etc.

A transgene is exogenous DNA that is integrated into the genome of a cell from which a transgenic animal develops and that remains in the genome of the mature animal, thereby directing the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal. As used herein, a "homologous recombinant animal" is a non-human animal, preferably a mammal, more preferably a mouse, in which an endogenous POLYX gene has been altered by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, *e.g.*, an embryonic cell of the animal, prior to development of the animal.

A transgenic animal of the invention can be created by introducing POLYX-encoding nucleic acid into the male pronuclei of a fertilized oocyte (*e.g.*, by micro-injection, retroviral infection) and allowing the oocyte to develop in a pseudopregnant female foster animal. The human POLYX cDNA sequences of SEQ ID NO:2*n*-1 (wherein n = 1 to 17), can be introduced as a transgene into the genome of a non-human animal. Alternatively, a non-human homologue of the human POLYX gene, such as a mouse POLYX gene, can be isolated based on hybridization to the human POLYX cDNA (described further *supra*) and used as a transgene. Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of expression of the transgene. A tissue-specific regulatory sequence(s) can be operably-linked to the POLYX transgene to direct expression of POLYX protein to particular cells. Methods for generating transgenic animals via embryo manipulation and micro-injection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866; 4,870,009; and 4,873,191; and Hogan, 1986. In: MANIPULATING THE MOUSE EMBRYO, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based

upon the presence of the POLYX transgene in its genome and/or expression of POLYX mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene-encoding POLYX protein can further be bred to other transgenic animals carrying other transgenes.

To create a homologous recombinant animal, a vector is prepared which contains at least a portion of a POLYX gene into which a deletion, addition or substitution has been introduced to thereby alter, *e.g.*, functionally disrupt, the POLYX gene. The POLYX gene can be a human gene (*e.g.*, the cDNA of SEQ ID NO:2*n*-1 (wherein n = 1 to 17)), but more preferably, is a non-human homologue of a human POLYX gene. For example, a mouse homologue of human POLYX gene of SEQ ID NO:2*n*-1 (wherein n = 1 to 17), can be used to construct a homologous recombination vector suitable for altering an endogenous POLYX gene in the mouse genome. In one embodiment, the vector is designed such that, upon homologous recombination, the endogenous POLYX gene is functionally disrupted (*i.e.*, no longer encodes a functional protein; also referred to as a "knock out" vector).

Alternatively, the vector can be designed such that, upon homologous recombination, the endogenous POLYX gene is mutated or otherwise altered but still encodes functional protein (*e.g.*, the upstream regulatory region can be altered to thereby alter the expression of the endogenous POLYX protein). In the homologous recombination vector, the altered portion of the POLYX gene is flanked at its 5'- and 3'-termini by additional nucleic acid of the POLYX gene to allow for homologous recombination to occur between the exogenous POLYX gene carried by the vector and an endogenous POLYX gene in an embryonic stem cell. The additional flanking POLYX nucleic acid is of sufficient length for successful homologous recombination with the endogenous gene. Typically, several kilobases (Kb) of flanking DNA (both at the 5'- and 3'-termini) are included in the vector. *See, e.g.,* Thomas, *et al.*, 1987. *Cell* 51: 503 for a description of homologous recombination vectors. The vector is ten introduced into an embryonic stem cell line (*e.g.*, by electroporation) and cells in which the introduced POLYX gene has homologously-recombined with the endogenous POLYX gene are selected. *See, e.g.*, Li, *et al.*, 1992. *Cell* 69: 915.

The selected cells are then micro-injected into a blastocyst of an animal (*e.g.*, a mouse) to form aggregation chimeras. *See, e.g.*, Bradley, 1987. In: TERATOCARCINOMAS AND EMBRYONIC STEM CELLS: A PRACTICAL APPROACH, Robertson, ed. IRL, Oxford, pp. 113-152. A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Progeny harboring the homologously-recombined DNA in their germ cells can be used to breed animals in which all cells of the animal contain the homologously-recombined DNA by germline transmission of the transgene. Methods for constructing homologous recombination vectors and homologous recombinant animals are described further in Bradley, 1991. *Curr. Opin. Biotechnol.* 2: 823-829; PCT International Publication Nos.: WO 90/11354; WO 91/01140; WO 92/0968; and WO 93/04169.

In another embodiment, transgenic non-human animals can be produced that contain selected systems that allow for regulated expression of the transgene. One example of such a system is the cre/loxP recombinase system of bacteriophage P1. For a description of the cre/loxP recombinase system, *See, e.g.*, Lakso, *et al.*, 1992. *Proc. Natl. Acad. Sci. USA* 89: 6232-6236. Another example of a recombinase system is the FLP recombinase system of *Saccharomyces cerevisiae*. *See,* O'Gorman, *et al.*, 1991. *Science* 251:1351-1355. If a cre/loxP recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the Cre recombinase and a selected protein are required. Such animals can be provided through the construction of "double" transgenic animals, *e.g.*, by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut, *et al.*, 1997. *Nature* 385: 810-813. In brief, a cell (*e.g.*, a somatic cell) from the transgenic animal can be isolated and induced to exit the growth cycle and enter G_0 phase. The quiescent cell can then be fused, *e.g.*, through the use of electrical pulses, to an enucleated oocyte from an animal of the same species from which the quiescent cell is isolated. The reconstructed oocyte is then cultured such that it develops to morula or blastocyte, and then transferred to pseudopregnant female foster animal. The offspring borne of this female foster animal will be a clone of the animal from which the cell (*e.g.*, the somatic cell) is isolated.

Pharmaceutical Compositions

The POLYX nucleic acid molecules, POLYX proteins, and anti-POLYX antibodies (also referred to herein as "active compounds") of the invention, and derivatives, fragments, analogs and homologs thereof, can be incorporated into pharmaceutical compositions suitable for administration. Such compositions typically comprise the nucleic acid molecule, protein, or antibody and a pharmaceutically-acceptable carrier. As used herein, "pharmaceutically-acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. Suitable carriers are described in the most recent edition of Remington's Pharmaceutical Sciences, a standard reference text in the field, which is incorporated herein by reference. Preferred examples of such carriers or diluents include, but are not limited to, water, saline, finger's solutions, dextrose solution, and 5% human serum albumin. Liposomes and other non-aqueous (*i.e.,* lipophilic) vehicles such as fixed oils may also be used. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, use thereof in the compositions is contemplated. Supplementary active compounds can also be incorporated into the compositions.

A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, *e.g.*, intravenous, intradermal, subcutaneous, oral (*e.g.*, inhalation), transdermal (*i.e.,* topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid (EDTA); buffers such as acetates, citrates or phosphates, and agents for the adjustment of tonicity such as sodium chloride or dextrose. The pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, N.J.) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringeability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as manitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions can be prepared by incorporating the active compound (*e.g.*, a POLYX protein or anti-POLYX antibody) in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle that contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, methods of preparation are vacuum drying and freeze-drying that yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For the purpose of oral therapeutic

administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash, wherein the compound in the fluid carrier is applied orally and swished and expectorated or swallowed. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser which contains a suitable propellant, *e.g.*, a gas such as carbon dioxide, or a nebulizer.

Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

The compounds can also be prepared in the form of suppositories (*e.g.*, with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of

such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

It is especially advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent on the unique characteristics of the active compound and the particular therapeutic effect to be achieved, and the limitations inherent in the art of compounding such an active compound for the treatment of individuals.

The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (*see, e.g.,* U.S. Patent No. 5,328,470) or by stereotactic injection (*see, e.g.*, Chen, *et al.*, 1994. *Proc. Natl. Acad. Sci. USA* 91: 3054-3057). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, *e.g.*, retroviral vectors, the pharmaceutical preparation can include one or more cells that produce the gene delivery system.

The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

Screening and Detection Methods

The nucleic acid molecules, proteins, protein homologues, and antibodies described herein can be used in one or more of the following methods: (*i*) screening assays; (*ii*) detection assays (*e.g.*, chromosomal mapping, cell and tissue typing, forensic biology), (*iii*) predictive medicine (*e.g.*, diagnostic assays, prognostic assays, monitoring clinical trials, and pharmacogenomics); and (*iv*) methods of treatment (*e.g.*, therapeutic and prophylactic).

The isolated nucleic acid molecules of the present invention can be used to express POLYX protein *(e.g.,* via a recombinant expression vector in a host cell in gene therapy applications), to detect POLYX mRNA (*e.g.,* in a biological sample) or a genetic lesion in an POLYX gene, and to modulate POLYX activity, as described further, *infra.* In addition, the POLYX proteins can be used to screen drugs or compounds that modulate the POLYX protein activity or expression as well as to treat disorders characterized by insufficient or excessive production of POLYX protein or production of POLYX protein forms that have decreased or aberrant activity compared to POLYX wild-type protein. In addition, the anti-POLYX antibodies of the present invention can be used to detect and isolate POLYX proteins and modulate POLYX activity.

The invention further pertains to novel agents identified by the screening assays described herein and uses thereof for treatments as described, supra.

Screening Assays

The invention provides a method (also referred to herein as a "screening assay") for identifying modulators, *i.e.*, candidate or test compounds or agents (*e.g.*, peptides, peptidomimetics, small molecules or other drugs) that bind to POLYX proteins or have a stimulatory or inhibitory effect on, *e.g.,* POLYX protein expression or POLYX protein activity. The invention also includes compounds identified in the screening assays described herein.

In one embodiment, the invention provides assays for screening candidate or test compounds which bind to or modulate the activity of the membrane-bound form of a POLYX

protein or polypeptide or biologically-active portion thereof. The test compounds of the invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the "one-bead one-compound" library method; and synthetic library methods using affinity chromatography selection. The biological library approach is limited to peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds. *See, e.g.,* Lam, 1997. *Anticancer Drug Design* 12: 145.

A "small molecule" as used herein, is meant to refer to a composition that has a molecular weight of less than about 5 kD and most preferably less than about 4 kD. Small molecules can be, *e.g.*, nucleic acids, peptides, polypeptides, peptidomimetics, carbohydrates, lipids or other organic or inorganic molecules. Libraries of chemical and/or biological mixtures, such as fungal, bacterial, or algal extracts, are known in the art and can be screened with any of the assays of the invention.

Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt, *et al.*, 1993. *Proc. Natl. Acad. Sci. U.S.A.* 90: 6909; Erb, *et al.*, 1994. *Proc. Natl. Acad. Sci. U.S.A.* 91: 11422; Zuckermann, *et al.*, 1994. *J. Med. Chem.* 37: 2678; Cho, *et al.*, 1993. *Science* 261: 1303; Carrell, *et al.*, 1994. *Angew. Chem. Int. Ed. Engl.* 33: 2059; Carell, *et al.*, 1994. *Angew. Chem. Int. Ed. Engl.* 33: 2061; and Gallop, *et al.*, 1994. *J. Med. Chem.* 37: 1233.

Libraries of compounds may be presented in solution (*e.g.*, Houghten, 1992. *Biotechniques* 13: 412-421), or on beads (Lam, 1991. *Nature* 354: 82-84), on chips (Fodor, 1993. *Nature* 364: 555-556), bacteria (Ladner, U.S. Patent No. 5,223,409), spores (Ladner, U.S. Patent 5,233,409), plasmids (Cull, *et al.*, 1992. *Proc. Natl. Acad. Sci. USA* 89: 1865-1869) or on phage (Scott and Smith, 1990. *Science* 249: 386-390; Devlin, 1990. *Science* 249: 404-406; Cwirla, *et al.*, 1990. *Proc. Natl. Acad. Sci. U.S.A.* 87: 6378-6382; Felici, 1991. *J. Mol. Biol.* 222: 301-310; Ladner, U.S. Patent No. 5,233,409.).

In one embodiment, an assay is a cell-based assay in which a cell which expresses a membrane-bound form of POLYX protein, or a biologically-active portion thereof, on the cell

surface is contacted with a test compound and the ability of the test compound to bind to a POLYX protein determined. The cell, for example, can of mammalian origin or a yeast cell. Determining the ability of the test compound to bind to the POLYX protein can be accomplished, for example, by coupling the test compound with a radioisotope or enzymatic label such that binding of the test compound to the POLYX protein or biologically-active portion thereof can be determined by detecting the labeled compound in a complex. For example, test compounds can be labeled with ^{125}I, ^{35}S, ^{14}C, or ^{3}H, either directly or indirectly, and the radioisotope detected by direct counting of radioemission or by scintillation counting. Alternatively, test compounds can be enzymatically-labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product. In one embodiment, the assay comprises contacting a cell which expresses a membrane-bound form of POLYX protein, or a biologically-active portion thereof, on the cell surface with a known compound which binds POLYX to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a POLYX protein, wherein determining the ability of the test compound to interact with a POLYX protein comprises determining the ability of the test compound to preferentially bind to POLYX protein or a biologically-active portion thereof as compared to the known compound.

In another embodiment, an assay is a cell-based assay comprising contacting a cell expressing a membrane-bound form of POLYX protein, or a biologically-active portion thereof, on the cell surface with a test compound and determining the ability of the test compound to modulate (*e.g.*, stimulate or inhibit) the activity of the POLYX protein or biologically-active portion thereof. Determining the ability of the test compound to modulate the activity of POLYX or a biologically-active portion thereof can be accomplished, for example, by determining the ability of the POLYX protein to bind to or interact with a POLYX target molecule. As used herein, a "target molecule" is a molecule with which a POLYX protein binds or interacts in nature, for example, a molecule on the surface of a cell which expresses a POLYX interacting protein, a molecule on the surface of a second cell, a molecule in the extracellular *milieu*, a molecule associated with the internal surface of a cell membrane or a cytoplasmic molecule. An POLYX target molecule can be a non-POLYX

molecule or a POLYX protein or polypeptide of the invention. In one embodiment, a POLYX target molecule is a component of a signal transduction pathway that facilitates transduction of an extracellular signal (*e.g.* a signal generated by binding of a compound to a membrane-bound POLYX molecule) through the cell membrane and into the cell. The target, for example, can be a second intercellular protein that has catalytic activity or a protein that facilitates the association of downstream signaling molecules with POLY.

Determining the ability of the POLYX protein to bind to or interact with a POLYX target molecule can be accomplished by one of the methods described above for determining direct binding. In one embodiment, determining the ability of the POLYX protein to bind to or interact with a POLYX target molecule can be accomplished by determining the activity of the target molecule. For example, the activity of the target molecule can be determined by detecting induction of a cellular second messenger of the target (*i.e.* intracellular Ca^{2+}, diacylglycerol, IP_3, etc.), detecting catalytic/enzymatic activity of the target an appropriate substrate, detecting the induction of a reporter gene (comprising a POLYX-responsive regulatory element operatively linked to a nucleic acid encoding a detectable marker, *e.g.*, luciferase), or detecting a cellular response, for example, cell survival, cellular differentiation, or cell proliferation.

In yet another embodiment, an assay of the invention is a cell-free assay comprising contacting a POLYX protein or biologically-active portion thereof with a test compound and determining the ability of the test compound to bind to the POLYX protein or biologically-active portion thereof. Binding of the test compound to the POLYX protein can be determined either directly or indirectly as described above. In one such embodiment, the assay comprises contacting the POLYX protein or biologically-active portion thereof with a known compound which binds POLYX to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a POLYX protein, wherein determining the ability of the test compound to interact with a POLYX protein comprises determining the ability of the test compound to preferentially bind to POLYX or biologically-active portion thereof as compared to the known compound.

In still another embodiment, an assay is a cell-free assay comprising contacting POLYX protein or biologically-active portion thereof with a test compound and determining the ability of the test compound to modulate (*e.g.* stimulate or inhibit) the activity of the POLYX protein or biologically-active portion thereof. Determining the ability of the test compound to modulate the activity of POLYX can be accomplished, for example, by determining the ability of the POLYX protein to bind to a POLYX target molecule by one of the methods described above for determining direct binding. In an alternative embodiment, determining the ability of the test compound to modulate the activity of POLYX protein can be accomplished by determining the ability of the POLYX protein further modulate a POLYX target molecule. For example, the catalytic/enzymatic activity of the target molecule on an appropriate substrate can be determined as described, *supra*.

In yet another embodiment, the cell-free assay comprises contacting the POLYX protein or biologically-active portion thereof with a known compound which binds POLYX protein to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a POLYX protein, wherein determining the ability of the test compound to interact with a POLYX protein comprises determining the ability of the POLYX protein to preferentially bind to or modulate the activity of a POLYX target molecule.

The cell-free assays of the invention are amenable to use of both the soluble form or the membrane-bound form of POLYX protein. In the case of cell-free assays comprising the membrane-bound form of POLYX protein, it may be desirable to utilize a solubilizing agent such that the membrane-bound form of POLYX protein is maintained in solution. Examples of such solubilizing agents include non-ionic detergents such as n-octylglucoside, n-dodecylglucoside, n-dodecylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton® X-100, Triton® X-114, Thesit®, Isotridecypoly(ethylene glycol ether)$_n$, N-dodecyl--N,N-dimethyl-3-ammonio-1-propane sulfonate, 3-(3-cholamidopropyl) dimethylamminiol-1-propane sulfonate (CHAPS), or 3-(3-cholamidopropyl)dimethylamminiol-2-hydroxy-1-propane sulfonate (CHAPSO).

In more than one embodiment of the above assay methods of the invention, it may be desirable to immobilize either POLYX protein or its target molecule to facilitate separation of complexed from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to POLYX protein, or interaction of POLYX protein with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtiter plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided that adds a domain that allows one or both of the proteins to be bound to a matrix. For example, GST-POLYX fusion proteins or GST-target fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtiter plates, that are then combined with the test compound or the test compound and either the non-adsorbed target protein or POLYX protein, and the mixture is incubated under conditions conducive to complex formation (*e.g.*, at physiological conditions for salt and pH). Following incubation, the beads or microtiter plate wells are washed to remove any unbound components, the matrix immobilized in the case of beads, complex determined either directly or indirectly, for example, as described, *supra*. Alternatively, the complexes can be dissociated from the matrix, and the level of POLYX protein binding or activity determined using standard techniques.

Other techniques for immobilizing proteins on matrices can also be used in the screening assays of the invention. For example, either the POLYX protein or its target molecule can be immobilized utilizing conjugation of biotin and streptavidin. Biotinylated POLYX protein or target molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques well-known within the art (*e.g.*, biotinylation kit, Pierce Chemicals, Rockford, Ill.), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with POLYX protein or target molecules, but which do not interfere with binding of the POLYX protein to its target molecule, can be derivatized to the wells of the plate, and unbound target or POLYX protein trapped in the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the POLYX protein or target

molecule, as well as enzyme-linked assays that rely on detecting an enzymatic activity associated with the POLYX protein or target molecule.

In another embodiment, modulators of POLYX protein expression are identified in a method wherein a cell is contacted with a candidate compound and the expression of POLYX mRNA or protein in the cell is determined. The level of expression of POLYX mRNA or protein in the presence of the candidate compound is compared to the level of expression of POLYX mRNA or protein in the absence of the candidate compound. The candidate compound can then be identified as a modulator of POLYX mRNA or protein expression based upon this comparison. For example, when expression of POLYX mRNA or protein is greater (*i.e.*, statistically significantly greater) in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of POLYX mRNA or protein expression. Alternatively, when expression of POLYX mRNA or protein is less (statistically significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of POLYX mRNA or protein expression. The level of POLYX mRNA or protein expression in the cells can be determined by methods described herein for detecting POLYX mRNA or protein.

In yet another aspect of the invention, the POLYX proteins can be used as "bait proteins" in a two-hybrid assay or three hybrid assay (*see, e.g.*, U.S. Patent No. 5,283,317; Zervos, *et al.*, 1993. *Cell* 72: 223-232; Madura, *et al.*, 1993. *J. Biol. Chem.* 268: 12046-12054; Bartel, *et al.*, 1993. *Biotechniques* 14: 920-924; Iwabuchi, *et al.*, 1993. *Oncogene* 8: 1693-1696; and Brent WO 94/10300), to identify other proteins that bind to or interact with POLYX ("POLYX-binding proteins" or "POLYX-bp") and modulate POLYX activity. Such POLYX-binding proteins are also likely to be involved in the propagation of signals by the POLYX proteins as, for example, upstream or downstream elements of the POLYX pathway.

The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for POLYX is fused to a gene encoding the DNA binding domain of a known transcription factor (*e.g.*, GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an

unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact, *in vivo*, forming a POLYX-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close POLYX imity. This POLYX imity allows transcription of a reporter gene (*e.g.*, LacZ) that is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene that encodes the protein which interacts with POLY.

The invention further pertains to novel agents identified by the aforementioned screening assays and uses thereof for treatments as described herein.

Detection Assays

Portions or fragments of the cDNA sequences identified herein (and the corresponding complete gene sequences) can be used in numerous ways as polynucleotide reagents. By way of example, and not of limitation, these sequences can be used to: (*i*) map their respective genes on a chromosome; and, thus, locate gene regions associated with genetic disease; (*ii*) identify an individual from a minute biological sample (tissue typing); and (*iii*) aid in forensic identification of a biological sample. Some of these applications are described in the subsections, *infra*.

Chromosome Mapping

Once the sequence (or a portion of the sequence) of a gene has been isolated, this sequence can be used to map the location of the gene on a chromosome. This process is called chromosome mapping. Accordingly, portions or fragments of the POLYX sequences shown in SEQ ID NO:2*n*-1 (wherein n = 1 to 17), or fragments or derivatives thereof, can be used to map the location of the POLYX genes, respectively, on a chromosome. The mapping of the POLYX sequences to chromosomes is an important first step in correlating these sequences with genes associated with disease.

Briefly, POLYX genes can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp in length) from the POLYX sequences. Computer analysis of the POLY, sequences can be used to rapidly select primers that do not span more than one exon in the genomic DNA, thus complicating the amplification process. These primers can then be used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the POLYX sequences will yield an amplified fragment.

Somatic cell hybrids are prepared by fusing somatic cells from different mammals (*e.g.*, human and mouse cells). As hybrids of human and mouse cells grow and divide, they gradually lose human chromosomes in random order, but retain the mouse chromosomes. By using media in which mouse cells cannot grow, because they lack a particular enzyme, but in which human cells can, the one human chromosome that contains the gene encoding the needed enzyme will be retained. By using various media, panels of hybrid cell lines can be established. Each cell line in a panel contains either a single human chromosome or a small number of human chromosomes, and a full set of mouse chromosomes, allowing easy mapping of individual genes to specific human chromosomes. *See, e.g.,* D'Eustachio, *et al.*, 1983. *Science* 220: 919-924. Somatic cell hybrids containing only fragments of human chromosomes can also be produced by using human chromosomes with translocations and deletions.

PCR mapping of somatic cell hybrids is a rapid procedure for assigning a particular sequence to a particular chromosome. Three or more sequences can be assigned per day using a single thermal cycler. Using the POLYX sequences to design oligonucleotide primers, sub-localization can be achieved with panels of fragments from specific chromosomes.

Fluorescence *in situ* hybridization (FISH) of a DNA sequence to a metaphase chromosomal spread can further be used to provide a precise chromosomal location in one step. Chromosome spreads can be made using cells whose division has been blocked in metaphase by a chemical like colcemid that disrupts the mitotic spindle. The chromosomes can be treated briefly with trypsin, and then stained with Giemsa. A pattern of light and dark bands develops on each chromosome, so that the chromosomes can be identified individually.

The FISH technique can be used with a DNA sequence as short as 500 or 600 bases. However, clones larger than 1,000 bases have a higher likelihood of binding to a unique chromosomal location with sufficient signal intensity for simple detection. Preferably 1,000 bases, and more preferably 2,000 bases, will suffice to get good results at a reasonable amount of time. For a review of this technique, *see,* Verma, *et al.*, HUMAN CHROMOSOMES: A MANUAL OF BASIC TECHNIQUES (Pergamon Press, NY 1988).

Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on that chromosome, or panels of reagents can be used for marking multiple sites and/or multiple chromosomes. Reagents corresponding to non-coding regions of the genes actually are preferred for mapping purposes. Coding sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. Such data are found, *e.g.,* in McKusick, MENDELIAN INHERITANCE IN MAN, available on-line through Johns Hopkins University Welch Medical Library). The relationship between genes and disease, mapped to the same chromosomal region, can then be identified through linkage analysis (co-inheritance of physically adjacent genes), described in, *e.g.,* Egeland, *et al.*, 1987. *Nature*, 325: 783-787.

Additionally, differences in the DNA sequences between individuals affected and unaffected with a disease associated with the POLYX gene, can be determined. If a mutation is observed in some or all of the affected individuals but not in any unaffected individuals, then the mutation is likely to be the causative agent of the particular disease. Comparison of affected and unaffected individuals generally involves first looking for structural alterations in the chromosomes, such as deletions or translocations that are visible from chromosome spreads or detectable using PCR based on that DNA sequence. Ultimately, complete sequencing of genes from several individuals can be performed to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

Tissue Typing

The POLYX sequences of the invention can also be used to identify individuals from minute biological samples. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identification. The sequences of the invention are useful as additional DNA markers for RFLP ("restriction fragment length polymorphisms," as described in U.S. Patent No. 5,272,057).

Furthermore, the sequences of the invention can be used to provide an alternative technique that determines the actual base-by-base DNA sequence of selected portions of an individual's genome. Thus, the POLYX sequences described herein can be used to prepare two PCR primers from the 5'- and 3'-termini of the sequences. These primers can then be used to amplify an individual's DNA and subsequently sequence it.

Panels of corresponding DNA sequences from individuals, prepared in this manner, can provide unique individual identifications, as each individual will have a unique set of such DNA sequences due to allelic differences. The sequences of the invention can be used to obtain such identification sequences from individuals and from tissue. The POLYX sequences of the invention uniquely represent portions of the human genome. Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the non-coding regions. It is estimated that allelic variation between individual humans occurs with a frequency of about once per each 500 bases. Much of the allelic variation is due to single nucleotide polymorphisms (SNPs), which include restriction fragment length polymorphisms (RFLPs).

Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes. Because greater numbers of polymorphisms occur in the non-coding regions, fewer sequences are necessary to differentiate individuals. The non-coding sequences can comfortably provide positive individual identification with a panel of perhaps 10 to 1,000 primers that each yield a non-coding amplified sequence of 100 bases. If predicted coding sequences, such as those in SEQ ID NO:2*n*-1 (wherein n = 1 to 17) are used, a more appropriate number of primers for positive individual identification would be 500-2,000.

Use of Partial POLYX Sequences in Forensic Biology

DNA-based identification techniques can also be used in forensic biology. Forensic biology is a scientific field employing genetic typing of biological evidence found at a crime scene as a means for positively identifying, *e.g.*, a perpetrator of a crime. To make such an identification, PCR technology can be used to amplify DNA sequences taken from very small biological samples such as tissues (*e.g.*, hair or skin, or body fluids, *e.g.*, blood, saliva, or semen found at a crime scene). The amplified sequence can then be compared to a standard, thereby allowing identification of the origin of the biological sample.

The sequences of the invention can be used to provide polynucleotide reagents, *e.g.*, PCR primers, targeted to specific loci in the human genome, that can enhance the reliability of DNA-based forensic identifications by, for example, providing another "identification marker" (*i.e.* another DNA sequence that is unique to a particular individual). As mentioned above, actual base sequence information can be used for identification as an accurate alternative to patterns formed by restriction enzyme generated fragments. Sequences targeted to non-coding regions of SEQ ID NO:2*n*-1 (where $n = 1$ to 17) are particularly appropriate for this use as greater numbers of polymorphisms occur in the non-coding regions, making it easier to differentiate individuals using this technique. Examples of polynucleotide reagents include the POLYX sequences or portions thereof, *e.g.*, fragments derived from the non-coding regions of one or more of SEQ ID NO:2*n*-1 (where $n = 1$ to 17), having a length of at least 20 bases, preferably at least 30 bases.

The POLYX sequences described herein can further be used to provide polynucleotide reagents, *e.g.*, labeled or label-able probes that can be used, for example, in an *in situ* hybridization technique, to identify a specific tissue (*e.g.*, brain tissue, etc). This can be very useful in cases where a forensic pathologist is presented with a tissue of unknown origin. Panels of such POLYX probes can be used to identify tissue by species and/or by organ type.

In a similar fashion, these reagents, *e.g.*, POLYX primers or probes can be used to screen tissue culture for contamination (*i.e.*, screen for the presence of a mixture of different types of cells in a culture).

Predictive Medicine

The invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, pharmacogenomics, and monitoring clinical trials are used for prognostic (predictive) purposes to thereby treat an individual prophylactically. Accordingly, one aspect of the invention relates to diagnostic assays for determining POLYX protein and/or nucleic acid expression as well as POLYX activity, in the context of a biological sample (*e.g.*, blood, serum, cells, tissue) to thereby determine whether an individual is afflicted with a disease or disorder, or is at risk of developing a disorder, associated with aberrant POLYX expression or activity. The invention also provides for prognostic (or predictive) assays for determining whether an individual is at risk of developing a disorder associated with POLYX protein, nucleic acid expression or activity. For example, mutations in a POLYX gene can be assayed in a biological sample. Such assays can be used for prognostic or predictive purpose to thereby prophylactically treat an individual prior to the onset of a disorder characterized by or associated with POLYX protein, nucleic acid expression, or biological activity.

Another aspect of the invention provides methods for determining POLYX protein, nucleic acid expression or activity in an individual to thereby select appropriate therapeutic or prophylactic agents for that individual (referred to herein as "pharmacogenomics"). Pharmacogenomics allows for the selection of agents (*e.g.*, drugs) for therapeutic or prophylactic treatment of an individual based on the genotype of the individual (*e.g.*, the genotype of the individual examined to determine the ability of the individual to respond to a particular agent.)

Yet another aspect of the invention pertains to monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of POLYX in clinical trials. These and other agents are described in further detail in the following sections.

Diagnostic Assays

An exemplary method for detecting the presence or absence of POLYX in a biological sample involves obtaining a biological sample from a test subject and contacting the biological sample with a compound or an agent capable of detecting POLYX protein or nucleic acid

(*e.g.*, mRNA, genomic DNA) that encodes POLYX protein such that the presence of POLYX is detected in the biological sample. An agent for detecting POLYX mRNA or genomic DNA is a labeled nucleic acid probe capable of hybridizing to POLYX mRNA or genomic DNA. The nucleic acid probe can be, for example, a full-length POLYX nucleic acid, such as the nucleic acid of SEQ ID NO:2*n*-1 (wherein n = 1 to 17), or a portion thereof, such as an oligonucleotide of at least 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to POLYX mRNA or genomic DNA. Other suitable probes for use in the diagnostic assays of the invention are described herein.

An agent for detecting POLYX protein is an antibody capable of binding to POLYX protein, preferably an antibody with a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (*e.g.*, F_{ab} or $F_{(ab)2}$) can be used. As utilized herein, the term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (*i.e.*, physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently-labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently-labeled streptavidin. As utilized herein, the term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. That is, the detection method of the invention can be used to detect POLYX mRNA, protein, or genomic DNA in a biological sample *in vitro* as well as *in vivo*. For example, *in vitro* techniques for detection of POLYX mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detection of POLYX protein include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations, and immunofluorescence. *In vitro* techniques for detection of POLYX genomic DNA include Southern hybridizations. Furthermore, *in vivo* techniques for detection of POLYX protein include introducing into a subject a labeled anti-POLYX antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

In one embodiment, the biological sample contains protein molecules from the test subject. Alternatively, the biological sample can contain mRNA molecules from the test subject or genomic DNA molecules from the test subject. A preferred biological sample is a peripheral blood leukocyte sample isolated by conventional means from a subject.

In another embodiment, the methods further involve obtaining a control biological sample from a control subject, contacting the control sample with a compound or agent capable of detecting POLYX protein, mRNA, or genomic DNA, such that the presence of POLYX protein, mRNA or genomic DNA is detected in the biological sample, and comparing the presence of POLYX protein, mRNA or genomic DNA in the control sample with the presence of POLYX protein, mRNA or genomic DNA in the test sample.

The invention also encompasses kits for detecting the presence of POLYX in a biological sample. For example, the kit can comprise: a labeled compound or agent capable of detecting POLYX protein or mRNA in a biological sample; means for determining the amount of POLYX in the sample; and means for comparing the amount of POLYX in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect POLYX protein or nucleic acid.

Prognostic Assays

The diagnostic methods described herein can furthermore be utilized to identify subjects having or at risk of developing a disease or disorder associated with aberrant POLYX expression or activity. For example, the assays described herein, such as the preceding diagnostic assays or the following assays, can be utilized to identify a subject having or at risk of developing a disorder associated with POLYX protein, nucleic acid expression or activity. Alternatively, the prognostic assays can be utilized to identify a subject having or at risk for developing a disease or disorder. Thus, the invention provides a method for identifying a disease or disorder associated with aberrant POLYX expression or activity in which a test sample is obtained from a subject and POLYX protein or nucleic acid (*e.g.*, mRNA, genomic DNA) is detected, wherein the presence of POLYX protein or nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant POLYX expression or activity. As used herein, a "test sample" refers to a biological sample obtained

from a subject of interest. For example, a test sample can be a biological fluid (*e.g.*, serum), cell sample, or tissue.

Furthermore, the prognostic assays described herein can be used to determine whether a subject can be administered an agent (*e.g.*, an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate) to treat a disease or disorder associated with aberrant POLYX expression or activity. For example, such methods can be used to determine whether a subject can be effectively treated with an agent for a disorder. Thus, the invention provides methods for determining whether a subject can be effectively treated with an agent for a disorder associated with aberrant POLYX expression or activity in which a test sample is obtained and POLYX protein or nucleic acid is detected (*e.g.*, wherein the presence of POLYX protein or nucleic acid is diagnostic for a subject that can be administered the agent to treat a disorder associated with aberrant POLYX expression or activity).

The methods of the invention can also be used to detect genetic lesions in a POLYX gene, thereby determining if a subject with the lesioned gene is at risk for a disorder characterized by aberrant cell proliferation and/or differentiation. In various embodiments, the methods include detecting, in a sample of cells from the subject, the presence or absence of a genetic lesion characterized by at least one of an alteration affecting the integrity of a gene encoding a POLYX-protein, or the mis-expression of the POLYX gene. For example, such genetic lesions can be detected by ascertaining the existence of at least one of: (*i*) a deletion of one or more nucleotides from a POLYX gene; (*ii*) an addition of one or more nucleotides to a POLYX gene; (*iii*) a substitution of one or more nucleotides of a POLYX gene, (*iv*) a chromosomal rearrangement of a POLYX gene; (*v*) an alteration in the level of a messenger RNA transcript of a POLYX gene; (*vi*) aberrant modification of a POLYX gene, such as of the methylation pattern of the genomic DNA; (*vii*) the presence of a non-wild-type splicing pattern of a messenger RNA transcript of a POLYX gene; (*viii*) a non-wild-type level of a POLYX protein, (*ix*) allelic loss of a POLYX gene; and (*x*) inappropriate post-translational modification of a POLYX protein. As described herein, there are a large number of assay techniques known in the art which can be used for detecting lesions in a POLYX gene. A preferred biological sample is a peripheral blood leukocyte sample isolated by conventional

means from a subject. However, any biological sample containing nucleated cells may be used, including, for example, buccal mucosal cells.

In certain embodiments, detection of the lesion involves the use of a probe/primer in a polymerase chain reaction (PCR) (*see, e.g.*, U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (*see, e.g.*, Landegran, *et al.*, 1988. *Science* 241: 1077-1080; and Nakazawa, *et al.*, 1994. *Proc. Natl. Acad. Sci. USA* 91: 360-364), the latter of which can be particularly useful for detecting point mutations in the POLYX-gene (*see,* Abravaya, *et al.*, 1995. *Nucl. Acids Res.* 23: 675-682). This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (*e.g.*, genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers that specifically hybridize to a POLYX gene under conditions such that hybridization and amplification of the POLYX gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

Alternative amplification methods include: self sustained sequence replication (*see,* Guatelli, *et al.*, 1990. *Proc. Natl. Acad. Sci. USA* 87: 1874-1878), transcriptional amplification system (*see,* Kwoh, *et al.*, 1989. *Proc. Natl. Acad. Sci. USA* 86: 1173-1177); Qβ Replicase (*see,* Lizardi, *et al*, 1988. *BioTechnology* 6: 1197), or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

In an alternative embodiment, mutations in a POLYX gene from a sample cell can be identified by alterations in restriction enzyme cleavage patterns. For example, sample and control DNA is isolated, amplified (optionally), digested with one or more restriction endonucleases, and fragment length sizes are determined by gel electrophoresis and compared. Differences in fragment length sizes between sample and control DNA indicates mutations in the sample DNA. Moreover, the use of sequence specific ribozymes (*see, e.g.*, U.S. Patent

No. 5,493,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

In other embodiments, genetic mutations in POLYX can be identified by hybridizing a sample and control nucleic acids, *e.g.*, DNA or RNA, to high-density arrays containing hundreds or thousands of oligonucleotides probes. *See, e.g.,* Cronin, *et al.*, 1996. *Human Mutation* 7: 244-255; Kozal, *et al.*, 1996. *Nat. Med.* 2: 753-759. For example, genetic mutations in POLYX can be identified in two dimensional arrays containing light-generated DNA probes as described in Cronin, *et al.*, *supra.* Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a sample and control to identify base changes between the sequences by making linear arrays of sequential overlapping probes. This step allows the identification of point mutations. This is followed by a second hybridization array that allows the characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

In yet another embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence the POLYX gene and detect mutations by comparing the sequence of the sample POLYX with the corresponding wild-type (control) sequence. Examples of sequencing reactions include those based on techniques developed by Maxim and Gilbert, 1977. *Proc. Natl. Acad. Sci. USA* 74: 560 or Sanger, 1977. *Proc. Natl. Acad. Sci. USA* 74: 5463. It is also contemplated that any of a variety of automated sequencing procedures can be utilized when performing the diagnostic assays (*see, e.g.,* Naeve, *et al.*, 1995. *BioTechniques* 19: 448), including sequencing by mass spectrometry (see, *e.g.*, PCT International Publication No. WO 94/16101; Cohen, *et al.*, 1996. *Adv. Chromatography* 36: 127-162; and Griffin, *et al.*, 1993. *Appl. Biochem. Biotechnol.* 38: 147-159).

Other methods for detecting mutations in the POLYX gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA heteroduplexes. *See, e.g.,* Myers, *et al.*, 1985. *Science* 230: 1242. In general, the art technique of "mismatch cleavage" starts by providing heteroduplexes of formed by

hybridizing (labeled) RNA or DNA containing the wild-type POLYX sequence with potentially mutant RNA or DNA obtained from a tissue sample. The double-stranded duplexes are treated with an agent that cleaves single-stranded regions of the duplex such as which will exist due to basepair mismatches between the control and sample strands. For instance, RNA/DNA duplexes can be treated with RNase and DNA/DNA hybrids treated with S_1 nuclease to enzymatically digesting the mismatched regions. In other embodiments, either DNA/DNA or RNA/DNA duplexes can be treated with hydroxylamine or osmium tetroxide and with piperidine in order to digest mismatched regions. After digestion of the mismatched regions, the resulting material is then separated by size on denaturing polyacrylamide gels to determine the site of mutation. *See, e.g.,* Cotton, *et al.,* 1988. *Proc. Natl. Acad. Sci. USA* 85: 4397; Saleeba, *et al.,* 1992. *Methods Enzymol.* 217: 286-295. In an embodiment, the control DNA or RNA can be labeled for detection.

In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize mismatched base pairs in double-stranded DNA (so called "DNA mismatch repair" enzymes) in defined systems for detecting and mapping point mutations in POLYX cDNAs obtained from samples of cells. For example, the mutY enzyme of *E. coli* cleaves A at G/A mismatches and the thymidine DNA glycosylase from HeLa cells cleaves T at G/T mismatches. *See, e.g.,* Hsu, *et al.,* 1994. *Carcinogenesis* 15: 1657-1662. According to an exemplary embodiment, a probe based on a POLYX sequence, *e.g.,* a wild-type POLYX sequence, is hybridized to a cDNA or other DNA product from a test cell(s). The duplex is treated with a DNA mismatch repair enzyme, and the cleavage products, if any, can be detected from electrophoresis protocols or the like. *See, e.g.,* U.S. Patent No. 5,459,039.

In other embodiments, alterations in electrophoretic mobility will be used to identify mutations in POLYX genes. For example, single strand conformation polymorphism (SSCP) may be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids. *See, e.g.,* Orita, *et al.,* 1989. *Proc. Natl. Acad. Sci. USA*: 86: 2766; Cotton, 1993. *Mutat. Res.* 285: 125-144; Hayashi, 1992. *Genet. Anal. Tech. Appl.* 9: 73-79. Single-stranded DNA fragments of sample and control POLYX nucleic acids will be denatured and allowed to renature. The secondary structure of single-stranded nucleic acids varies according to sequence, the resulting alteration in electrophoretic mobility enables the

detection of even a single base change. The DNA fragments may be labeled or detected with labeled probes. The sensitivity of the assay may be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In one embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility. *See, e.g.,* Keen, *et al.*, 1991. *Trends Genet.* 7: 5.

In yet another embodiment, the movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE). *See, e.g.,* Myers, *et al.*, 1985. *Nature* 313: 495. When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a GC clamp of apPOLYXimately 40 bp of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a denaturing gradient to identify differences in the mobility of control and sample DNA. *See, e.g.,* Rosenbaum and Reissner, 1987. *Biophys. Chem.* 265: 12753.

Examples of other techniques for detecting point mutations include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or selective primer extension. For example, oligonucleotide primers may be prepared in which the known mutation is placed centrally and then hybridized to target DNA under conditions that permit hybridization only if a perfect match is found. *See, e.g.,* Saiki, *et al.*, 1986. *Nature* 324: 163; Saiki, *et al.*, 1989. *Proc. Natl. Acad. Sci. USA* 86: 6230. Such allele specific oligonucleotides are hybridized to PCR amplified target DNA or a number of different mutations when the oligonucleotides are attached to the hybridizing membrane and hybridized with labeled target DNA.

Alternatively, allele specific amplification technology that depends on selective PCR amplification may be used in conjunction with the instant invention. Oligonucleotides used as primers for specific amplification may carry the mutation of interest in the center of the molecule (so that amplification depends on differential hybridization; *see, e.g.,* Gibbs, *et al.*, 1989. *Nucl. Acids Res.* 17: 2437-2448) or at the extreme 3'-terminus of one primer where, under appropriate conditions, mismatch can prevent, or reduce polymerase extension (*see, e.g.,*

Prossner, 1993. *Tibtech.* 11: 238). In addition it may be desirable to introduce a novel restriction site in the region of the mutation to create cleavage-based detection. *See, e.g.,* Gasparini, *et al.,* 1992. *Mol. Cell Probes* 6: 1. It is anticipated that in certain embodiments amplification may also be performed using *Taq* ligase for amplification. *See, e.g.,* Barany, 1991. *Proc. Natl. Acad. Sci. USA* 88: 189. In such cases, ligation will occur only if there is a perfect match at the 3'-terminus of the 5' sequence, making it possible to detect the presence of a known mutation at a specific site by looking for the presence or absence of amplification.

The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one probe nucleic acid or antibody reagent described herein, which may be conveniently used, *e.g.,* in clinical settings to diagnose patients exhibiting symptoms or family history of a disease or illness involving a POLYX gene.

Furthermore, any cell type or tissue, preferably peripheral blood leukocytes, in which POLYX is expressed may be utilized in the prognostic assays described herein. However, any biological sample containing nucleated cells may be used, including, for example, buccal mucosal cells.

Pharmacogenomics

Agents, or modulators that have a stimulatory or inhibitory effect on POLYX activity (*e.g.,* POLYX gene expression), as identified by a screening assay described herein can be administered to individuals to treat (prophylactically or therapeutically) disorders (*e.g.,* cancer or immune disorders associated with aberrant POLYX activity. In conjunction with such treatment, the pharmacogenomics (*i.e.,* the study of the relationship between an individual's genotype and that individual's response to a foreign compound or drug) of the individual may be considered. Differences in metabolism of therapeutics can lead to severe toxicity or therapeutic failure by altering the relation between dose and blood concentration of the pharmacologically active drug. Thus, the pharmacogenomics of the individual permits the selection of effective agents (*e.g.,* drugs) for prophylactic or therapeutic treatments based on a consideration of the individual's genotype. Such pharmacogenomics can further be used to determine appropriate dosages and therapeutic regimens. Accordingly, the activity of POLYX

protein, expression of POLYX nucleic acid, or mutation content of POLYX genes in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual.

Pharmacogenomics deals with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See *e.g.*, Eichelbaum, 1996. *Clin. Exp. Pharmacol. Physiol.* 23: 983-985; Linder, 1997. *Clin. Chem.*, 43: 254-266. In general, two types of pharmacogenetic conditions can be differentiated. Genetic conditions transmitted as a single factor altering the way drugs act on the body (altered drug action) or genetic conditions transmitted as single factors altering the way the body acts on drugs (altered drug metabolism). These pharmacogenetic conditions can occur either as rare defects or as polymorphisms. For example, glucose-6-phosphate dehydrogenase (G6PD) deficiency is a common inherited enzymopathy in which the main clinical complication is hemolysis after ingestion of oxidant drugs (anti-malarials, sulfonamides, analgesics, nitrofurans) and consumption of fava beans.

As an illustrative embodiment, the activity of drug metabolizing enzymes is a major determinant of both the intensity and duration of drug action. The discovery of genetic polymorphisms of drug metabolizing enzymes (*e.g.*, N-acetyltransferase 2 (NAT 2) and cytochrome P450 enzymes CYP2D6 and CYP2C19) has provided an explanation as to why some patients do not obtain the expected drug effects or show exaggerated drug response and serious toxicity after taking the standard and safe dose of a drug. These polymorphisms are expressed in two phenotypes in the population, the extensive metabolizer (EM) and poor metabolizer (PM). The prevalence of PM is different among different populations. For example, the gene coding for CYP2D6 is highly polymorphic and several mutations have been identified in PM, which all lead to the absence of functional CYP2D6. Poor metabolizers of CYP2D6 and CYP2C19 quite frequently experience exaggerated drug response and side effects when they receive standard doses. If a metabolite is the active therapeutic moiety, PM show no therapeutic response, as demonstrated for the analgesic effect of codeine mediated by its CYP2D6-formed metabolite morphine. At the other extreme are the so called ultra-rapid metabolizers who do not respond to standard doses. Recently, the molecular basis of ultra-rapid metabolism has been identified to be due to CYP2D6 gene amplification.

Thus, the activity of POLYX protein, expression of POLYX nucleic acid, or mutation content of POLYX genes in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual. In addition, pharmacogenetic studies can be used to apply genotyping of polymorphic alleles encoding drug-metabolizing enzymes to the identification of an individual's drug responsiveness phenotype. This knowledge, when applied to dosing or drug selection, can avoid adverse reactions or therapeutic failure and thus enhance therapeutic or prophylactic efficiency when treating a subject with a POLYX modulator, such as a modulator identified by one of the exemplary screening assays described herein.

Monitoring of Effects During Clinical Trials

Monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of POLYX (*e.g.*, the ability to modulate aberrant cell proliferation and/or differentiation) can be applied not only in basic drug screening, but also in clinical trials. For example, the effectiveness of an agent determined by a screening assay as described herein to increase POLYX gene expression, protein levels, or upregulate POLYX activity, can be monitored in clinical trails of subjects exhibiting decreased POLYX gene expression, protein levels, or downregulated POLYX activity. Alternatively, the effectiveness of an agent determined by a screening assay to decrease POLYX gene expression, protein levels, or downregulate POLYX activity, can be monitored in clinical trails of subjects exhibiting increased POLYX gene expression, protein levels, or upregulated POLYX activity. In such clinical trials, the expression or activity of POLYX and, preferably, other genes that have been implicated in, for example, a cellular proliferation or immune disorder can be used as a "read out" or markers of the immune responsiveness of a particular cell.

By way of example, and not of limitation, genes, including POLY, that are modulated in cells by treatment with an agent (*e.g.*, compound, drug or small molecule) that modulates POLYX activity (*e.g.*, identified in a screening assay as described herein) can be identified. Thus, to study the effect of agents on cellular proliferation disorders, for example, in a clinical trial, cells can be isolated and RNA prepared and analyzed for the levels of expression of POLYX and other genes implicated in the disorder. The levels of gene expression (*i.e.*, a gene

expression pattern) can be quantified by Northern blot analysis or RT-PCR, as described herein, or alternatively by measuring the amount of protein produced, by one of the methods as described herein, or by measuring the levels of activity of POLYX or other genes. In this manner, the gene expression pattern can serve as a marker, indicative of the physiological response of the cells to the agent. Accordingly, this response state may be determined before, and at various points during, treatment of the individual with the agent.

In one embodiment, the invention provides a method for monitoring the effectiveness of treatment of a subject with an agent (*e.g.*, an agonist, antagonist, protein, peptide, peptidomimetic, nucleic acid, small molecule, or other drug candidate identified by the screening assays described herein) comprising the steps of (*i*) obtaining a pre-administration sample from a subject prior to administration of the agent; (*ii*) detecting the level of expression of a POLYX protein, mRNA, or genomic DNA in the pre-administration sample; (*iii*) obtaining one or more post-administration samples from the subject; (*iv*) detecting the level of expression or activity of the POLYX protein, mRNA, or genomic DNA in the post-administration samples; (*v*) comparing the level of expression or activity of the POLYX protein, mRNA, or genomic DNA in the pre-administration sample with the POLYX protein, mRNA, or genomic DNA in the post administration sample or samples; and (*vi*) altering the administration of the agent to the subject accordingly. For example, increased administration of the agent may be desirable to increase the expression or activity of POLYX to higher levels than detected, *i.e.*, to increase the effectiveness of the agent. Alternatively, decreased administration of the agent may be desirable to decrease expression or activity of POLYX to lower levels than detected, *i.e.*, to decrease the effectiveness of the agent.

Methods of Treatment

The invention provides for both prophylactic and therapeutic methods of treating a subject at risk of (or susceptible to) a disorder or having a disorder associated with aberrant POLYX expression or activity. These methods of treatment will be discussed more fully, *infra.*

Disease and Disorders

Diseases and disorders that are characterized by increased (relative to a subject not suffering from the disease or disorder) levels or biological activity may be treated with Therapeutics that antagonize (*i.e.*, reduce or inhibit) activity. Therapeutics that antagonize activity may be administered in a therapeutic or prophylactic manner. Therapeutics that may be utilized include, but are not limited to: (*i*) an aforementioned peptide, or analogs, derivatives, fragments or homologs thereof; (*ii*) antibodies to an aforementioned peptide; (*iii*) nucleic acids encoding an aforementioned peptide; (*iv*) administration of antisense nucleic acid and nucleic acids that are "dysfunctional" (*i.e.*, due to a heterologous insertion within the coding sequences of coding sequences to an aforementioned peptide) that are utilized to "knockout" endoggenous function of an aforementioned peptide by homologous recombination (*see, e.g.,* Capecchi, 1989. *Science* 244: 1288-1292); or (*v*) modulators (*i.e.*, inhibitors, agonists and antagonists, including additional peptide mimetic of the invention or antibodies specific to a peptide of the invention) that alter the interaction between an aforementioned peptide and its binding partner.

Diseases and disorders that are characterized by decreased (relative to a subject not suffering from the disease or disorder) levels or biological activity may be treated with Therapeutics that increase (*i.e.*, are agonists to) activity. Therapeutics that upregulate activity may be administered in a therapeutic or prophylactic manner. Therapeutics that may be utilized include, but are not limited to, an aforementioned peptide, or analogs, derivatives, fragments or homologs thereof; or an agonist that increases bioavailability.

Increased or decreased levels can be readily detected by quantifying peptide and/or RNA, by obtaining a patient tissue sample (*e.g.,* from biopsy tissue) and assaying it *in vitro* for RNA or peptide levels, structure and/or activity of the expressed peptides (or mRNAs of an aforementioned peptide). Methods that are well-known within the art include, but are not limited to, immunoassays (*e.g.,* by Western blot analysis, immunoprecipitation followed by sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis, immunocytochemistry, etc.) and/or hybridization assays to detect expression of mRNAs (*e.g.,* Northern assays, dot blots, *in situ* hybridization, and the like).

Prophylactic Methods

In one aspect, the invention provides a method for preventing, in a subject, a disease or condition associated with an aberrant POLYX expression or activity, by administering to the subject an agent that modulates POLYX expression or at least one POLYX activity. Subjects at risk for a disease that is caused or contributed to by aberrant POLYX expression or activity can be identified by, for example, any or a combination of diagnostic or prognostic assays as described herein. Administration of a prophylactic agent can occur prior to the manifestation of symptoms characteristic of the POLYX aberrancy, such that a disease or disorder is prevented or, alternatively, delayed in its progression. Depending upon the type of POLYX aberrancy, for example, a POLYX agonist or POLYX antagonist agent can be used for treating the subject. The appropriate agent can be determined based on screening assays described herein.

Therapeutic Methods

Another aspect of the invention pertains to methods of modulating POLYX expression or activity for therapeutic purposes. The modulatory method of the invention involves contacting a cell with an agent that modulates one or more of the activities of POLYX protein activity associated with the cell. An agent that modulates POLYX protein activity can be an agent as described herein, such as a nucleic acid or a protein, a naturally-occurring cognate ligand of a POLYX protein, a peptide, a POLYX peptidomimetic, or other small molecule. In one embodiment, the agent stimulates one or more POLYX protein activity. Examples of such stimulatory agents include active POLYX protein and a nucleic acid molecule encoding POLYX that has been introduced into the cell. In another embodiment, the agent inhibits one or more POLYX protein activity. Examples of such inhibitory agents include antisense POLYX nucleic acid molecules and anti-POLYX antibodies. These modulatory methods can be performed *in vitro* (*e.g.*, by culturing the cell with the agent) or, alternatively, *in vivo* (*e.g.*, by administering the agent to a subject). As such, the invention provides methods of treating an individual afflicted with a disease or disorder characterized by aberrant expression or activity of a POLYX protein or nucleic acid molecule. In one embodiment, the method involves administering an agent (*e.g.*, an agent identified by a screening assay described

herein), or combination of agents that modulates (*e.g.*, up-regulates or down-regulates) POLYX expression or activity. In another embodiment, the method involves administering a POLYX protein or nucleic acid molecule as therapy to compensate for reduced or aberrant POLYX expression or activity.

Stimulation of POLYX activity is desirable in situations in which POLYX is abnormally downregulated and/or in which increased POLYX activity is likely to have a beneficial effect. One example of such a situation is where a subject has a disorder characterized by aberrant cell proliferation and/or differentiation (*e.g.*, cancer or immune associated disorders). Another example of such a situation is where the subject has a gestational disease (*e.g.*, pre-clampsia).

Determination of the Biological Effect of the Therapeutic

In various embodiments of the invention, suitable *in vitro* or *in vivo* assays are performed to determine the effect of a specific Therapeutic and whether its administration is indicated for treatment of the affected tissue.

In various specific embodiments, *in vitro* assays may be performed with representative cells of the type(s) involved in the patient's disorder, to determine if a given Therapeutic exerts the desired effect upon the cell type(s). Compounds for use in therapy may be tested in suitable animal model systems including, but not limited to rats, mice, chicken, cows, monkeys, rabbits, and the like, prior to testing in human subjects. Similarly, for *in vivo* testing, any of the animal model system known in the art may be used prior to administration to human subjects.

Prophylactic and Therapeutic Uses of the Compositions of the Invention

The POLYX nucleic acids and proteins of the invention may be useful in a variety of potential prophylactic and therapeutic applications. By way of a non-limiting example, a cDNA encoding the POLYX protein of the invention may be useful in gene therapy, and the protein may be useful when administered to a subject in need thereof.

Both the novel nucleic acids encoding the POLYX proteins, and the POLYX proteins of the invention, or fragments thereof, may also be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed. These materials are further useful in the generation of antibodies which immunospecifically-bind to the novel substances of the invention for use in therapeutic or diagnostic methods.

The invention will be further illustrated in the following non-limiting examples.

Example 1: IDENTIFICATION OF POLYX NUCLEIC ACIDS

TblastN using CuraGen Corporation's sequence file for polypeptides or homologs was run against the Genomic Daily Files made available by GenBank or from files downloaded from the individual sequencing centers. Exons were predicted by homology and the intron/exon boundaries were determined using standard genetic rules. Exons were further selected and refined by means of similarity determination using multiple BLAST (for example, tBlastN, BlastX, and BlastN) searches, and, in some instances, GeneScan and Grail. Expressed sequences from both public and proprietary databases were also added when available to further define and complete the gene sequence. The DNA sequence was then manually corrected for apparent inconsistencies thereby obtaining the sequences encoding the full-length protein.

EXAMPLE 2: CLONING OF POLY3

The sequence of Acc. No. CG54683-02 (POLY3) was derived by laboratory cloning of cDNA fragments, by *in silico* prediction of the sequence. cDNA fragments covering either the full length of the DNA sequence, or part of the sequence, or both, were cloned. In silico prediction was based on sequences available in Curagen's proprietary sequence databases or in the public human sequence databases, and provided either the full length DNA sequence, or some portion thereof.

The cDNA coding for the CG54683-02 sequence was cloned by the polymerase chain reaction (PCR) using the primers:

5' TTGGAAGAGATGGTCCTGGCTTTC 3' (SEQ ID NO: 43) and

5' TTCATAGGATTCTCAGCTGTGTGAGTG 3' (SEQ ID NO:44).

Primers were designed based on in silico predictions of the full length or some portion (one or more exons) of the cDNA/protein sequence of the invention. These primers were used to amplify a cDNA from a pool containing expressed human sequences derived from the following tissues: adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, brain - thalamus, brain -whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea and uterus.

Multiple clones were sequenced and these fragments were assembled together, sometimes including public human sequences, using bioinformatic programs to produce a consensus sequence for each assembly. Each assembly is included in CuraGen Corporation's database. Sequences were included as components for assembly when the extent of identity with another component was at least 95% over 50 bp. Each assembly represents a gene or portion thereof and includes information on variants, such as splice forms single nucleotide polymorphisms (SNPs), insertions, deletions and other sequence variations.

EXAMPLE 3: IDENTIFICATION OF SINGLE NUCLEOTIDE POLYMORPHISMS IN POLY3 NUCLEIC ACID SEQUENCES

Variant sequences are also included in this application. A variant sequence can include a single nucleotide polymorphism (SNP). A SNP can, in some instances, be referred to as a "cSNP" to denote that the nucleotide sequence containing the SNP originates as a cDNA. A SNP can arise in several ways. For example, a SNP may be due to a substitution of one nucleotide for another at the polymorphic site. Such a substitution can be either a transition or a transversion. A SNP can also arise from a deletion of a nucleotide or an insertion of a nucleotide, relative to a reference allele. In this case, the polymorphic site is a site at which one allele bears a gap with respect to a particular nucleotide in another allele. SNPs occurring within genes may result in an alteration of the amino acid encoded by the gene at the position

of the SNP. Intragenic SNPs may also be silent, when a codon including a SNP encodes the same amino acid as a result of the redundancy of the genetic code. SNPs occurring outside the region of a gene, or in an intron within a gene, do not result in changes in any amino acid sequence of a protein but may result in altered regulation of the expression pattern. Examples include alteration in temporal expression, physiological response regulation, cell type expression regulation, intensity of expression, and stability of transcribed message.

SeqCalling assemblies produced by the exon linking process were selected and extended using the following criteria. Genomic clones having regions with 98% identity to all or part of the initial or extended sequence were identified by BLASTN searches using the relevant sequence to query human genomic databases. The genomic clones that resulted were selected for further analysis because this identity indicates that these clones contain the genomic locus for these SeqCalling assemblies. These sequences were analyzed for putative coding regions as well as for similarity to the known DNA and protein sequences. Programs used for these analyses include Grail, Genscan, BLAST, HMMER, FASTA, Hybrid and other relevant programs.

Some additional genomic regions may have also been identified because selected SeqCalling assemblies map to those regions. Such SeqCalling sequences may have overlapped with regions defined by homology or exon prediction. They may also be included because the location of the fragment was in the vicinity of genomic regions identified by similarity or exon prediction that had been included in the original predicted sequence. The sequence so identified was manually assembled and then may have been extended using one or more additional sequences taken from CuraGen Corporation's human SeqCalling database. SeqCalling fragments suitable for inclusion were identified by the CuraTools™ program SeqExtend or by identifying SeqCalling fragments mapping to the appropriate regions of the genomic clones analyzed. Such sequences were included in the derivation of Acc. No. CG54683-02 (POLY3) only when the extent of identity in the overlap region with one or more SeqCalling assemblies 160154242 was high. The extent of identity may be, for example, about 90% or higher, preferably about 95% or higher, and even more preferably close to or equal to 100%. When necessary, the process to identify and analyze SeqCalling fragments and genomic clones was reiterated to derive the full length sequence.

The regions defined by the procedures described above were then manually integrated and corrected for apparent inconsistencies that may have arisen, for example, from miscalled bases in the original fragments or from discrepancies between predicted exon junctions, EST locations and regions of sequence similarity, to derive the final sequence disclosed herein. When necessary, the process to identify and analyze SeqCalling assemblies and genomic clones was reiterated to derive the full length sequence.

The following public components were thus included in the invention: : gb_AC024892.10 HTG Homo sapiens|Homo sapiens chromosome 3 clone RP11-214N20, WORKING DRAFT SEQUENCE, 14 unordered pieces, 155257 bp. In addition, the following Curagen Corporation SeqCalling Assembly ID's were also included in the invention: 160154242.

EXAMPLE 4. IDENTIFICATION OF POLY4.

The sequence of POLY4 (Acc. No. CG54683-03) was derived by laboratory cloning of cDNA fragments, by *in silico* prediction of the sequence. cDNA fragments covering either the full length of the DNA sequence, or part of the sequence, or both, were cloned. In silico prediction was based on sequences available in Curagen's proprietary sequence databases or in the public human sequence databases, and provided either the full length DNA sequence, or some portion thereof.

The cDNA coding for the CG54683-03 sequence was cloned by the polymerase chain reaction (PCR) using the primers:

5' TTGGAAGAGATGGTCCTGGCTTTC 3' (SEQ ID NO:45) and

5' TTCATAGGATTCTCAGCTGTGTGAGTG 3' (SEQ ID NO:46).

Primers were designed based on in silico predictions of the full length or some portion (one or more exons) of the cDNA/protein sequence of the invention. These primers were used to amplify a cDNA from a pool containing expressed human sequences derived from the following tissues: adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, brain - thalamus, brain -whole, fetal brain, fetal kidney,

fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea and uterus.

Multiple clones were sequenced and these fragments were assembled together, sometimes including public human sequences, using bioinformatic programs to produce a consensus sequence for each assembly. Each assembly is included in CuraGen Corporation's database. Sequences were included as components for assembly when the extent of identity with another component was at least 95% over 50 bp. Each assembly represents a gene or portion thereof and includes information on variants, such as splice forms single nucleotide polymorphisms (SNPs), insertions, deletions and other sequence variations.

Variant sequences are also included in this application. A variant sequence can include a single nucleotide polymorphism (SNP). A SNP can, in some instances, be referred to as a "cSNP" to denote that the nucleotide sequence containing the SNP originates as a cDNA. A SNP can arise in several ways. For example, a SNP may be due to a substitution of one nucleotide for another at the polymorphic site. Such a substitution can be either a transition or a transversion. A SNP can also arise from a deletion of a nucleotide or an insertion of a nucleotide, relative to a reference allele. In this case, the polymorphic site is a site at which one allele bears a gap with respect to a particular nucleotide in another allele. SNPs occurring within genes may result in an alteration of the amino acid encoded by the gene at the position of the SNP. Intragenic SNPs may also be silent, when a codon including a SNP encodes the same amino acid as a result of the redundancy of the genetic code. SNPs occurring outside the region of a gene, or in an intron within a gene, do not result in changes in any amino acid sequence of a protein but may result in altered regulation of the expression pattern. Examples include alteration in temporal expression, physiological response regulation, cell type expression regulation, intensity of expression, and stability of transcribed message.

SeqCalling assemblies produced by the exon linking process were selected and extended using the following criteria. Genomic clones having regions with 98% identity to all or part of the initial or extended sequence were identified by BLASTN searches using the relevant sequence to query human genomic databases. The genomic clones that resulted were

selected for further analysis because this identity indicates that these clones contain the genomic locus for these SeqCalling assemblies. These sequences were analyzed for putative coding regions as well as for similarity to the known DNA and protein sequences. Programs used for these analyses include Grail, Genscan, BLAST, HMMER, FASTA, Hybrid and other relevant programs.

Some additional genomic regions may have also been identified because selected SeqCalling assemblies map to those regions. Such SeqCalling sequences may have overlapped with regions defined by homology or exon prediction. They may also be included because the location of the fragment was in the vicinity of genomic regions identified by similarity or exon prediction that had been included in the original predicted sequence. The sequence so identified was manually assembled and then may have been extended using one or more additional sequences taken from CuraGen Corporation's human SeqCalling database. SeqCalling fragments suitable for inclusion were identified by the CuraTools™ program SeqExtend or by identifying SeqCalling fragments mapping to the appropriate regions of the genomic clones analyzed. Such sequences were included in the derivation of POLY4 only when the extent of identity in the overlap region with one or more SeqCalling assemblies 160154242 was high. The extent of identity may be, for example, about 90% or higher, preferably about 95% or higher, and even more preferably close to or equal to 100%. When necessary, the process to identify and analyze SeqCalling fragments and genomic clones was reiterated to derive the full length sequence.

The regions defined by the procedures described above were then manually integrated and corrected for apparent inconsistencies that may have arisen, for example, from miscalled bases in the original fragments or from discrepancies between predicted exon junctions, EST locations and regions of sequence similarity, to derive the final sequence disclosed herein. When necessary, the process to identify and analyze SeqCalling assemblies and genomic clones was reiterated to derive the full length sequence. The following public components were thus included in the invention: gb_AC024892.10 HTG Homo sapiens|Homo sapiens chromosome 3 clone RP11-214N20, WORKING DRAFT SEQUENCE, 14 unordered pieces, 155257 bp. In addition, the following Curagen Corporation SeqCalling Assembly ID's were also included in the invention: 160154242.

EXAMPLE 5. - Quantitative expression analysis of clones in various cells and tissues

The quantitative expression of various clones was assessed using microtiter plates containing RNA samples from a variety of normal and pathology-derived cells, cell lines and tissues using real time quantitative PCR (RTQ PCR; TAQMAN®). RTQ PCR was performed in a Perkin-Elmer Biosystems ABI PRISM® 7700 Sequence Detection System. Various collections of samples are assembled on the plates, and referred to as Panel 1 (containing cells and cell lines from normal and cancer sources), Panel 2 (containing samples derived from tissues, in particular from surgical samples, from normal and cancer sources), Panel 3 (containing samples derived from a wide variety of cancer sources), Panel 4 (containing cells and cell lines from normal cells and cells related to inflammatory conditions) and Panel CNSD.01 (containing samples from normal and diseased brains).

First, the RNA samples were normalized to constitutively expressed genes such as b-lactin and GAPDH. RNA (~50 ng total or ~1 ng polyA+) was converted to cDNA using the TAQMAN® Reverse Transcription Reagents Kit (PE Biosystems, Foster City, CA; Catalog No. N808-0234) and random hexamers according to the manufacturer's protocol. Reactions were performed in 20 ul and incubated for 30 min. at 480C. cDNA (5 ul) was then transferred to a separate plate for the TAQMAN® reaction using b-actin and GAPDH TAQMAN® Assay Reagents (PE Biosystems; Catalog Nos. 4310881E and 4310884E, respectively) and TAQMAN® universal PCR Master Mix (PE Biosystems; Catalog No. 4304447) according to the manufacturer's protocol. Reactions were performed in 25 ul using the following parameters: 2 min. at 500C; 10 min. at 950C; 15 sec. at 950C/1 min. at 600C (40 cycles). Results were recorded as CT values (cycle at which a given sample crosses a threshold level of fluorescence) using a log scale, with the difference in RNA concentration between a given sample and the sample with the lowest CT value being represented as 2 to the power of delta CT. The percent relative expression is then obtained by taking the reciprocal of this RNA difference and multiplying by 100. The average CT values obtained for ß-actin and GAPDH were used to normalize RNA samples. The RNA sample generating the highest CT value required no further diluting, while all other samples were diluted relative to this sample according to their b-actin /GAPDH average CT values.

Normalized RNA (5 ul) was converted to cDNA and analyzed via TAQMAN® using One Step RT-PCR Master Mix Reagents (PE Biosystems; Catalog No. 4309169) and gene-specific primers according to the manufacturer's instructions. Probes and primers were designed for each assay according to Perkin Elmer Biosystem's Primer Express Software package (version I for Apple Computer's Macintosh Power PC) or a similar algorithm using the target sequence as input. Default settings were used for reaction conditions and the following parameters were set before selecting primers: primer concentration = 250 nM, primer melting temperature (Tm) range = 58°-60° C, primer optimal Tm = 59° C, maximum primer difference = 2° C, probe does not have 5' G, probe Tm must be 10° C greater than primer Tm, amplicon size 75 bp to 100 bp. The probes and primers selected (see below) were synthesized by Synthegen (Houston, TX, USA). Probes were double purified by HPLC to remove uncoupled dye and evaluated by mass spectroscopy to verify coupling of reporter and quencher dyes to the 5' and 3' ends of the probe, respectively. Their final concentrations were: forward and reverse primers, 900 nM each, and probe, 200nM.PCR conditions: Normalized RNA from each tissue and each cell line was spotted in each well of a 96 well PCR plate (Perkin Elmer Biosystems). PCR cocktails including two probes (a probe specific for the target clone and another gene-specific probe multiplexed with the target probe) were set up using 1X TaqManÔ PCR Master Mix for the PE Biosystems 7700, with 5 mM MgCl2, dNTPs (dA, G, C, U at 1:1:1:2 ratios), 0.25 U/ml AmpliTaq GoldÔ (PE Biosystems), and 0.4 U/ml RNase inhibitor, and 0.25 U/ml reverse transcriptase. Reverse transcription was performed at 48° C for 30 minutes followed by amplification/PCR cycles as follows: 95° C 10 min, then 40 cycles of 95° C for 15 seconds, 60° C for 1 minute.

In the results for Panel 1, the following abbreviations are used:

ca. = carcinoma,

* = established from metastasis,

met = metastasis,

s cell var= small cell variant,

non-s = non-sm =non-small,

squam = squamous,

pl. eff = pl effusion = pleural effusion,

glio = glioma,

astro = astrocytoma, and

neuro = neuroblastoma.

Panel 2

The plates for Panel 2 generally include 2 control wells and 94 test samples composed of RNA or cDNA isolated from human tissue procured by surgeons working in close cooperation with the National Cancer Institute's Cooperative Human Tissue Network (CHTN) or the National Disease Research Initiative (NDRI). The tissues are derived from human malignancies and in cases where indicated many malignant tissues have "matched margins" obtained from noncancerous tissue just adjacent to the tumor. These are termed normal adjacent tissues and are denoted "NAT" in the results below. The tumor tissue and the "matched margins" are evaluated by two independent pathologists (the surgical pathologists and again by a pathologists at NDRI or CHTN). This analysis provides a gross histopathological assessment of tumor differentiation grade. Moreover, most samples include the original surgical pathology report that provides information regarding the clinical stage of the patient. These matched margins are taken from the tissue surrounding (i.e. immediately proximal) to the zone of surgery (designated "NAT", for normal adjacent tissue, in Table RR). In addition, RNA and cDNA samples were obtained from various human tissues derived from autopsies performed on elderly people or sudden death victims (accidents, etc.). These tissues were ascertained to be free of disease and were purchased from various commercial sources such as Clontech (Palo Alto, CA), Research Genetics, and Invitrogen.

RNA integrity from all samples is controlled for quality by visual assessment of agarose gel electropherograms using 28S and 18S ribosomal RNA staining intensity ratio as a guide (2:1 to 2.5:1 28s:18s) and the absence of low molecular weight RNAs that would be indicative of degradation products. Samples are controlled against genomic DNA contamination by RTQ PCR reactions run in the absence of reverse transcriptase using probe and primer sets designed to amplify across the span of a single exon.

Panel 3D

The plates of Panel 3D are comprised of 94 cDNA samples and two control samples. Specifically, 92 of these samples are derived from cultured human cancer cell lines, 2 samples of human primary cerebellar tissue and 2 controls. The human cell lines are generally obtained from ATCC (American Type Culture Collection), NCI or the German tumor cell bank and fall into the following tissue groups: Squamous cell carcinoma of the tongue, breast cancer, prostate cancer, melanoma, epidermoid carcinoma, sarcomas, bladder carcinomas, pancreatic cancers, kidney cancers, leukemias/lymphomas, ovarian/uterine/cervical, gastric, colon, lung and CNS cancer cell lines. In addition, there are two independent samples of cerebellum. These cells are all cultured under standard recommended conditions and RNA extracted using the standard procedures. The cell lines in panel 3D and 1.3D are of the most common cell lines used in the scientific literature.

RNA integrity from all samples is controlled for quality by visual assessment of agarose gel electropherograms using 28S and 18S ribosomal RNA staining intensity ratio as a guide (2:1 to 2.5:1 28s:18s) and the absence of low molecular weight RNAs that would be indicative of degradation products. Samples are controlled against genomic DNA contamination by RTQ PCR reactions run in the absence of reverse transcriptase using probe and primer sets designed to amplify across the span of a single exon.

Panel 4

Panel 4 includes samples on a 96 well plate (2 control wells, 94 test samples) composed of RNA (Panel 4r) or cDNA (Panel 4d) isolated from various human cell lines or tissues related to inflammatory conditions. Total RNA from control normal tissues such as

colon and lung (Stratagene ,La Jolla, CA) and thymus and kidney (Clontech) were employed. Total RNA from liver tissue from cirrhosis patients and kidney from lupus patients was obtained from BioChain (Biochain Institute, Inc., Hayward, CA). Intestinal tissue for RNA preparation from patients diagnosed as having Crohn's disease and ulcerative colitis was obtained from the National Disease Research Interchange (NDRI) (Philadelphia, PA).

Astrocytes, lung fibroblasts, dermal fibroblasts, coronary artery smooth muscle cells, small airway epithelium, bronchial epithelium, microvascular dermal endothelial cells, microvascular lung endothelial cells, human pulmonary aortic endothelial cells, human umbilical vein endothelial cells were all purchased from Clonetics (Walkersville, MD) and grown in the media supplied for these cell types by Clonetics. These primary cell types were activated with various cytokines or combinations of cytokines for 6 and/or 12-14 hours, as indicated. The following cytokines were used; IL-1 beta at approximately 1-5 ng/ml, TNF alpha at approximately 5-10 ng/ml, IFN gamma at approximately 20-50 ng/ml, IL-4 at approximately 5-10 ng/ml, IL-9 at approximately 5-10 ng/ml, IL-13 at approximately 5-10 ng/ml. Endothelial cells were sometimes starved for various times by culture in the basal media from Clonetics with 0.1% serum.

Mononuclear cells were prepared from blood of employees at CuraGen Corporation, using Ficoll. LAK cells were prepared from these cells by culture in DMEM 5% FCS (Hyclone), 100 mM non essential amino acids (Gibco/Life Technologies, Rockville, MD), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5 x 10-5 M (Gibco), and 10 mM Hepes (Gibco) and Interleukin 2 for 4-6 days. Cells were then either activated with 10-20 ng/ml PMA and 1-2 mg/ml ionomycin, IL-12 at 5-10 ng/ml, IFN gamma at 20-50 ng/ml and IL-18 at 5-10 ng/ml for 6 hours. In some cases, mononuclear cells were cultured for 4-5 days in DMEM 5% FCS (Hyclone), 100 mM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5 x 10-5 M (Gibco), and 10 mM Hepes (Gibco) with PHA (phytohemagglutinin) or PWM (pokeweed mitogen) at approximately 5 mg/ml. Samples were taken at 24, 48 and 72 hours for RNA preparation. MLR (mixed lymphocyte reaction) samples were obtained by taking blood from two donors, isolating the mononuclear cells using Ficoll and mixing the isolated mononuclear cells 1:1 at a final concentration of approximately 2x106 cells/ml in DMEM 5% FCS (Hyclone), 100 mM non essential amino acids (Gibco), 1

mM sodium pyruvate (Gibco), mercaptoethanol (5.5 x 10-5 M) (Gibco), and 10 mM Hepes (Gibco). The MLR was cultured and samples taken at various time points ranging from 1- 7 days for RNA preparation.

Monocytes were isolated from mononuclear cells using CD14 Miltenyi Beads, +ve VS selection columns and a Vario Magnet according to the manufacturer's instructions. Monocytes were differentiated into dendritic cells by culture in DMEM 5% fetal calf serum (FCS) (Hyclone, Logan, UT), 100 mM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5 x 10-5 M (Gibco), and 10 mM Hepes (Gibco), 50 ng/ml GMCSF and 5 ng/ml IL-4 for 5-7 days. Macrophages were prepared by culture of monocytes for 5-7 days in DMEM 5% FCS (Hyclone), 100 mM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5 x 10-5 M (Gibco), 10 mM Hepes (Gibco) and 10% AB Human Serum or MCSF at approximately 50 ng/ml. Monocytes, macrophages and dendritic cells were stimulated for 6 and 12-14 hours with lipopolysaccharide (LPS) at 100 ng/ml. Dendritic cells were also stimulated with anti-CD40 monoclonal antibody (Pharmingen) at 10 mg/ml for 6 and 12-14 hours.

CD4 lymphocytes, CD8 lymphocytes and NK cells were also isolated from mononuclear cells using CD4, CD8 and CD56 Miltenyi beads, positive VS selection columns and a Vario Magnet according to the manufacturer's instructions. CD45RA and CD45RO CD4 lymphocytes were isolated by depleting mononuclear cells of CD8, CD56, CD14 and CD19 cells using CD8, CD56, CD14 and CD19 Miltenyi beads and positive selection. Then CD45RO beads were used to isolate the CD45RO CD4 lymphocytes with the remaining cells being CD45RA CD4 lymphocytes. CD45RA CD4, CD45RO CD4 and CD8 lymphocytes were placed in DMEM 5% FCS (Hyclone), 100 mM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5 x 10-5 M (Gibco), and 10 mM Hepes (Gibco) and plated at 106 cells/ml onto Falcon 6 well tissue culture plates that had been coated overnight with 0.5 mg/ml anti-CD28 (Pharmingen) and 3 ug/ml anti-CD3 (OKT3, ATCC) in PBS. After 6 and 24 hours, the cells were harvested for RNA preparation. To prepare chronically activated CD8 lymphocytes, we activated the isolated CD8 lymphocytes for 4 days on anti-CD28 and anti-CD3 coated plates and then harvested the cells and expanded them in DMEM 5% FCS (Hyclone), 100 mM non essential amino acids (Gibco), 1 mM sodium

pyruvate (Gibco), mercaptoethanol 5.5 x 10-5 M (Gibco), and 10 mM Hepes (Gibco) and IL-2. The expanded CD8 cells were then activated again with plate bound anti-CD3 and anti-CD28 for 4 days and expanded as before. RNA was isolated 6 and 24 hours after the second activation and after 4 days of the second expansion culture. The isolated NK cells were cultured in DMEM 5% FCS (Hyclone), 100 mM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5 x 10-5 M (Gibco), and 10 mM Hepes (Gibco) and IL-2 for 4-6 days before RNA was prepared.

To obtain B cells, tonsils were procured from NDRI. The tonsil was cut up with sterile dissecting scissors and then passed through a sieve. Tonsil cells were then spun down and resupended at 106 cells/ml in DMEM 5% FCS (Hyclone), 100 mM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5 x 10-5 M (Gibco), and 10 mM Hepes (Gibco). To activate the cells, we used PWM at 5 mg/ml or anti-CD40 (Pharmingen) at approximately 10 mg/ml and IL-4 at 5-10 ng/ml. Cells were harvested for RNA preparation at 24,48 and 72 hours.

To prepare the primary and secondary Th1/Th2 and Tr1 cells, six-well Falcon plates were coated overnight with 10 µg/ml anti-CD28 (Pharmingen) and 2 µg/ml OKT3 (ATCC), and then washed twice with PBS. Umbilical cord blood CD4 lymphocytes (Poietic Systems, German Town, MD) were cultured at 105-106 cells/ml in DMEM 5% FCS (Hyclone), 100 mM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5 x 10-5 M (Gibco), 10 mM Hepes (Gibco) and IL-2 (4 ng/ml). IL-12 (5 ng/ml) and anti-IL4 (1 □g/ml) were used to direct to Th1, while IL-4 (5 ng/ml) and anti-IFN gamma (1 □g/ml) were used to direct to Th2 and IL-10 at 5 ng/ml was used to direct to Tr1. After 4-5 days, the activated Th1, Th2 and Tr1 lymphocytes were washed once in DMEM and expanded for 4-7 days in DMEM 5% FCS (Hyclone), 100 mM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5 x 10-5 M (Gibco), 10 mM Hepes (Gibco) and IL-2 (1 ng/ml). Following this, the activated Th1, Th2 and Tr1 lymphocytes were re-stimulated for 5 days with anti-CD28/OKT3 and cytokines as described above, but with the addition of anti-CD95L (1 □g/ml) to prevent apoptosis. After 4-5 days, the Th1, Th2 and Tr1 lymphocytes were washed and then expanded again with IL-2 for 4-7 days. Activated Th1 and Th2 lymphocytes were maintained in this way for a maximum of three cycles. RNA was prepared

from primary and secondary Th1, Th2 and Tr1 after 6 and 24 hours following the second and third activations with plate bound anti-CD3 and anti-CD28 mAbs and 4 days into the second and third expansion cultures in Interleukin 2.

The following leukocyte cells lines were obtained from the ATCC: Ramos, EOL-1, KU-812. EOL cells were further differentiated by culture in 0.1 mM dbcAMP at 5 x105 cells/ml for 8 days, changing the media every 3 days and adjusting the cell concentration to 5 x105 cells/ml. For the culture of these cells, we used DMEM or RPMI (as recommended by the ATCC), with the addition of 5% FCS (Hyclone), 100 mM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5 x 10-5 M (Gibco), 10 mM Hepes (Gibco). RNA was either prepared from resting cells or cells activated with PMA at 10 ng/ml and ionomycin at 1 mg/ml for 6 and 14 hours. Keratinocyte line CCD106 and an airway epithelial tumor line NCI-H292 were also obtained from the ATCC. Both were cultured in DMEM 5% FCS (Hyclone), 100 mM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5 x 10-5 M (Gibco), and 10 mM Hepes (Gibco). CCD1106 cells were activated for 6 and 14 hours with approximately 5 ng/ml TNF alpha and 1 ng/ml IL-1 beta, while NCI-H292 cells were activated for 6 and 14 hours with the following cytokines: 5 ng/ml IL-4, 5 ng/ml IL-9, 5 ng/ml IL-13 and 25 ng/ml IFN gamma.

For these cell lines and blood cells, RNA was prepared by lysing approximately 107 cells/ml using Trizol (Gibco BRL). Briefly, 1/10 volume of bromochloropropane (Molecular Research Corporation) was added to the RNA sample, vortexed and after 10 minutes at room temperature, the tubes were spun at 14,000 rpm in a Sorvall SS34 rotor. The aqueous phase was removed and placed in a 15 ml Falcon Tube. An equal volume of isopropanol was added and left at –20 degrees C overnight. The precipitated RNA was spun down at 9,000 rpm for 15 min in a Sorvall SS34 rotor and washed in 70% ethanol. The pellet was redissolved in 300 ml of RNAse-free water and 35 ml buffer (Promega) 5 ml DTT, 7 ml RNAsin and 8 ml DNAse were added. The tube was incubated at 37 degrees C for 30 minutes to remove contaminating genomic DNA, extracted once with phenol chloroform and re-precipitated with 1/10 volume of 3 M sodium acetate and 2 volumes of 100% ethanol. The RNA was spun down and placed in RNAse free water. RNA was stored at –80 degrees C.

Panel CNSD.01

The plates for Panel CNSD.01 include two control wells and 94 test samples comprised of cDNA isolated from postmortem human brain tissue obtained from the Harvard Brain Tissue Resource Center. Brains are removed from calvaria of donors between 4 and 24 hours after death, sectioned by neuroanatomists, and frozen at -80oC in liquid nitrogen vapor. All brains are sectioned and examined by neuropathologists to confirm diagnoses with clear associated neuropathology.

Disease diagnoses are taken from patient records. The panel contains two brains from each of the following diagnoses: Alzheimer's disease, Parkinson's disease, Huntington's disease, Progressive Supernuclear Palsy, Depression, and "Normal controls". Within each of these brains, the following regions are represented: cingulate gyrus, temporal pole, globus palladus, substantia nigra, Brodman Area 4 (primary motor strip), Brodman Area 7 (parietal cortex), Brodman Area 9 (prefrontal cortex), and Brodman area 17 (occipital cortex). Not all brain regions are represented in all cases; e.g., Huntington's disease is characterized in part by neurodegeneration in the globus palladus, thus this region is impossible to obtain from confirmed Huntington's cases. Likewise Parkinson's disease is characterized by degeneration of the substantia nigra making this region more difficult to obtain. Normal control brains were examined for neuropathology and found to be free of any pathology consistent with neurodegeneration.

RNA integrity from all samples is controlled for quality by visual assessment of agarose gel electropherograms using 28S and 18S ribosomal RNA staining intensity ratio as a guide (2:1 to 2.5:1 28s:18s) and the absence of low molecular weight RNAs that would be indicative of degradation products. Samples are controlled against genomic DNA contamination by RTQ PCR reactions run in the absence of reverse transcriptase using probe and primer sets designed to amplify across the span of a single exon.

In the labels employed to identify tissues in the CNS panel, the following abbreviations are used:

PSP = Progressive supranuclear palsy

Sub Nigra = Substantia nigra

Glob Palladus= Globus palladus

Temp Pole = Temporal pole

Cing Gyr = Cingulate gyrus

BA 4 = Brodman Area 4

EXAMPLE 5A. EXPRESSION OF POLY5, POLY6 AND POLY7 NUCLEIC ACIDS IN CELLS AND TISSUES

GM-Z97832B:

Expression of gene GM-Z97832B was assessed using the primer-probe sets Ag192, Ag339, and Ag2890 (identical sequence as Ag1391), described in Tables 19, 20, and 21. Results of the RTQ-PCR runs are shown in Tables 22, 23, 24, 25, 26, and 27.

Table 19. Probe Name Ag192

Primers	Sequences	TM	Length	Start Position
Forward	5'-TGTGCCGAGGGCAACG-3'		16	208
Probe	FAM-5'-TAGCTGCCCATCATGTTGACACAGCTCT-3'-TAMRA		28	236
Reverse	5'-AGAAGCCTTCCCGGCAGT-3'		18	272

Table 20. Probe Name Ag339

Primers	Sequences	TM	Length	Start Position
Forward	5'-CTTGTGGGCCAGCTCATAATC-3'		21	1741
Probe	FAM-5'-ACCGCTTCCTGCTGCGCCTG-3'-TAMRA		20	1712
Reverse	5'-GATGCTCAGAAAGTCCATCAACC-3'		23	1686

Table 21. Probe Name Ag2890/Ag1391

Primers	Sequences	TM	Length	Start Position
Forward	5'-CTTATGAGACCTGCCAGACCTA-3'	58.5	22	2585
Probe	FAM-5'-CTTCACTGCCCGTTCCAGGAAGCT-3'-TAMRA	70.7	24	2622
Reverse	5'-CTCGCTTGTCTTGAAGTTGATC-3'	59.1	22	2649

Table 22. Panel 1

Tissue Name	Relative Expression(%) 1tm383f_ ag192	1tm356f_ ag192	Relative Expression(%) 1tm375f_ ag339	1tm568t_ ag339
Endothelial cells	32.1	0.0	0.6	1.4
Endothelial cells (treated)	15.4	0.0	0.0	0.2
Pancreas	19.5	0.2	2.9	1.4
Pancreatic ca. CAPAN 2	94.0	0.0	0.2	1.2
Adipose	23.3	0.3	17.3	15.5
Adrenal gland	38.2	4.0	11.8	8.1
Thyroid	52.8	28.9	59.0	35.6
Salivary gland	15.9	1.5	3.5	3.2
Pituitary gland	16.8	1.2	1.7	2.1
Brain (fetal)	37.4	1.2	4.6	2.5
Brain (whole)	23.2	0.7	15.6	0.9
Brain (amygdala)	33.0	0.7	1.9	2.7
Brain (cerebellum)	18.9	12.9	55.9	3.5
Brain (hippocampus)	23.0	0.3	1.5	2.2
Brain (substantia nigra)	17.8	0.5	5.1	2.6
Brain (thalamus)	26.8	1.0	5.9	3.0
Brain (hypothalamus)	42.9	8.1	10.2	8.3
Spinal cord	8.8	0.7	4.6	2.0
CNS ca. (glio/astro) U87-MG	18.4	0.0	0.6	1.5
CNS ca. (glio/astro) U-118-MG	15.3	2.6	7.5	2.4
CNS ca. (astro) SW1783	6.9	0.0	0.3	0.9
CNS ca.* (neuro; met) SK-N-AS	18.4	1.2	2.2	2.2
CNS ca. (astro) SF-539	10.9	0.9	1.7	2.0
CNS ca. (astro) SNB-75	**100.0**	**100.0**	**100.0**	50.7
CNS ca. (glio) SNB-19	15.3	6.5	19.9	4.5
CNS ca. (glio) U251	31.4	5.0	14.7	7.0
CNS ca. (glio) SF-295	11.2	1.9	5.4	3.1
Heart	22.8	3.1	16.0	13.2
Skeletal muscle	18.7	0.0	0.1	3.0

Bone marrow	16.5	0.0	0.5	3.9
Thymus	2.0	0.9	19.5	3.0
Spleen	14.4	0.2	0.0	1.2
Lymph node	20.2	0.6	1.5	1.3
Colon (ascending)	9.6	0.0	3.3	2.4
Stomach	2.2	0.2	18.4	1.9
Small intestine	16.8	0.0	1.9	1.8
Colon ca. SW480	7.7	0.0	0.0	1.6
Colon ca.* (SW480 met)SW620	12.7	0.2	0.9	2.1
Colon ca. HT29	8.7	0.0	0.0	0.6
Colon ca. HCT-116	14.6	0.2	1.4	3.5
Colon ca. CaCo-2	9.3	0.7	4.9	3.4
Colon ca. HCT-15	7.5	0.2	3.9	4.9
Colon ca. HCC-2998	9.7	0.2	5.1	3.1
Gastric ca.* (liver met) NCI-N87	5.7	0.0	2.9	1.7
Bladder	43.5	8.7	38.7	11.3
Trachea	15.5	2.0	6.6	1.7
Kidney	21.5	3.1	10.0	13.4
Kidney (fetal)	52.1	34.2	51.0	49.0
Renal ca. 786-0	6.9	32.8	5.3	2.1
Renal ca. A498	7.9	0.0	1.6	0.7
Renal ca. RXF 393	18.2	0.4	3.3	2.2
Renal ca. ACHN	4.8	0.0	1.6	1.6
Renal ca. UO-31	10.7	0.1	0.4	1.5
Renal ca. TK-10	9.5	1.7	3.8	3.7
Liver	21.3	0.0	0.0	1.2
Liver (fetal)	17.4	0.0	0.3	1.6
Liver ca. (hepatoblast) HepG2	66.9	0.3	2.1	9.5
Lung	2.9	1.0	13.0	1.0
Lung (fetal)	17.3	1.1	3.0	0.5
Lung ca. (small cell) LX-1	8.8	0.0	0.9	2.9
Lung ca. (small cell) NCI-H69	10.7	0.0	5.7	2.9
Lung ca. (s.cell var.) SHP-77	33.4	3.7	11.2	1.9
Lung ca. (large cell)NCI-H460	13.0	0.5	6.6	6.4
Lung ca. (non-sm. cell) A549	9.7	1.0	5.1	4.4
Lung ca. (non-s.cell) NCI-H23	18.4	4.4	7.8	14.2
Lung ca (non-s.cell) HOP-62	8.1	1.5	8.4	6.7
Lung ca. (non-s.cl) NCI-H522	38.7	21.5	44.1	23.8
Lung ca. (squam.) SW 900	8.3	1.0	8.5	5.2
Lung ca. (squam.) NCI-H596	27.0	4.4	10.4	8.0
Mammary gland	23.3	4.1	17.2	7.0

Breast ca.* (pl. effusion) MCF-7	9.5	0.0	3.8	2.3
Breast ca.* (pl.ef) MDA-MB-231	21.2	0.5	3.6	5.0
Breast ca.* (pl. effusion) T47D	11.9	0.4	8.1	4.0
Breast ca. BT-549	16.8	0.5	8.0	1.7
Breast ca. MDA-N	12.9	0.0	3.3	2.4
Ovary	2.1	0.0	10.1	3.7
Ovarian ca. OVCAR-3	14.4	0.5	3.5	2.9
Ovarian ca. OVCAR-4	12.9	5.6	5.3	5.1
Ovarian ca. OVCAR-5	8.4	2.1	3.6	3.4
Ovarian ca. OVCAR-8	11.3	7.2	12.8	18.7
Ovarian ca. IGROV-1	8.8	0.0	0.2	1.4
Ovarian ca.* (ascites) SK-OV-3	21.3	1.2	6.2	4.2
Uterus	22.5	4.8	12.1	8.1
Placenta	48.0	0.2	5.3	2.9
Prostate	18.7	2.7	6.9	3.6
Prostate ca.* (bone met)PC-3	27.4	24.1	49.7	**100.0**
Testis	3.1	17.8	64.6	5.8
Melanoma Hs688(A).T	25.0	0.0	3.7	1.7
Melanoma* (met) Hs688(B).T	69.3	51.4	97.3	92.0
Melanoma UACC-62	17.9	0.0	0.0	2.0
Melanoma M14	6.9	0.0	5.2	2.0
Melanoma LOX IMVI	5.1	0.3	20.2	1.7
Melanoma* (met) SK-MEL-5	17.6	0.4	3.5	3.3
Melanoma SK-MEL-28	20.7	0.9	5.4	4.2

Table 23. Panel 1.1

Tissue Name	Relative Expression(%) 1.1tm720f_ ag192	Tissue Name	Relative Expression(%) 1.1tm720f_ ag192
Adipose	0.5	Renal ca. TK-10	7.1
Adrenal gland	7.0	Renal ca. UO-31	1.6
Bladder	8.6	Renal ca. RXF 393	0.7
Brain (amygdala)	0.0	Liver	1.2
Brain (cerebellum)	4.7	Liver (fetal)	0.0
Brain (hippocampus)	1.6	Liver ca. (hepatoblast) HepG2	1.4
Brain (substantia nigra)	13.1	Lung	0.2
Brain (thalamus)	3.8	Lung (fetal)	3.3
Cerebral Cortex	2.1	Lung ca (non-s.cell) HOP-62	49.3
Brain (fetal)	2.4	Lung ca. (large cell)NCI-H460	5.8
Brain (whole)	2.4	Lung ca. (non-s.cell) NCI-H23	12.2

CNS ca. (glio/astro) U-118-MG	7.1	Lung ca. (non-s.cl) NCI-H522	77.9
CNS ca. (astro) SF-539	0.0	Lung ca. (non-sm. cell) A549	5.8
CNS ca. (astro) SNB-75	78.5	Lung ca. (s.cell var.) SHP-77	0.8
CNS ca. (astro) SW1783	0.2	Lung ca. (small cell) LX-1	2.5
CNS ca. (glio) U251	19.6	Lung ca. (small cell) NCI-H69	5.3
CNS ca. (glio) SF-295	9.2	Lung ca. (squam.) SW 900	3.8
CNS ca. (glio) SNB-19	15.9	Lung ca. (squam.) NCI-H596	14.4
CNS ca. (glio/astro) U87-MG	1.9	Lymph node	1.9
CNS ca.* (neuro; met) SK-N-AS	7.0	Spleen	0.0
Mammary gland	0.3	Thymus	0.3
Breast ca. BT-549	1.5	Ovary	5.6
Breast ca. MDA-N	1.6	Ovarian ca. IGROV-1	3.1
Breast ca.* (pl. effusion) T47D	5.3	Ovarian ca. OVCAR-3	8.6
Breast ca.* (pl. effusion) MCF-7	0.9	Ovarian ca. OVCAR-4	27.2
Breast ca.* (pl.ef) MDA-MB-231	5.5	Ovarian ca. OVCAR-5	9.9
Small intestine	2.2	Ovarian ca. OVCAR-8	17.6
Colorectal	0.9	Ovarian ca.* (ascites) SK-OV-3	11.3
Colon ca. HT29	0.3	Pancreas	10.0
Colon ca. CaCo-2	0.0	Pancreatic ca. CAPAN 2	0.0
Colon ca. HCT-15	2.9	Pituitary gland	10.9
Colon ca. HCT-116	1.5	Placenta	1.6
Colon ca. HCC-2998	3.8	Prostate	2.6
Colon ca. SW480	0.3	Prostate ca.* (bone met)PC-3	80.7
Colon ca.* (SW480 met)SW620	3.1	Salivary gland	8.1
Stomach	5.6	Trachea	3.9
Gastric ca.* (liver met) NCI-N87	1.8	Spinal cord	5.8
Heart	39.2	Testis	3.3
Fetal Skeletal	1.0	Thyroid	**100.0**
Skeletal muscle	7.4	Uterus	2.1
Endothelial cells	7.9	Melanoma M14	1.4
Endothelial cells (treated)	12.4	Melanoma LOX IMVI	0.5
Kidney	27.4	Melanoma UACC-62	0.2
Kidney (fetal)	37.1	Melanoma SK-MEL-28	3.5
Renal ca. 786-0	2.0	Melanoma* (met) SK-MEL-5	1.2
Renal ca. A498	0.6	Melanoma Hs688(A).T	0.3
Renal ca. ACHN	2.7	Melanoma* (met) Hs688(B).T	66.0

Table 24. Panel 1.3D

Tissue Name	Relative Expression(%) 1.3dx4tm_5798f_ag2890_a1	Tissue Name	Relative Expression(%) 1.3dx4tm_5798f_ag2890_a1
Liver adenocarcinoma	2.3	Kidney (fetal)	22.0
Pancreas	0.5	Renal ca. 786-0	1.9
Pancreatic ca. CAPAN 2	0.2	Renal ca. A498	7.7
Adrenal gland	1.7	Renal ca. RXF 393	3.3
Thyroid	15.7	Renal ca. ACHN	1.7
Salivary gland	0.6	Renal ca. UO-31	0.3
Pituitary gland	1.5	Renal ca. TK-10	2.0
Brain (fetal)	4.8	Liver	0.3
Brain (whole)	1.9	Liver (fetal)	0.3
Brain (amygdala)	1.0	Liver ca. (hepatoblast) HepG2	0.9
Brain (cerebellum)	3.6	Lung	0.7
Brain (hippocampus)	1.4	Lung (fetal)	1.4
Brain (substantia nigra)	2.1	Lung ca. (small cell) LX-1	1.5
Brain (thalamus)	1.8	Lung ca. (small cell) NCI-H69	2.2
Cerebral Cortex	1.8	Lung ca. (s.cell var.) SHP-77	7.3
Spinal cord	1.9	Lung ca. (large cell)NCI-H460	0.8
CNS ca. (glio/astro) U87-MG	1.3	Lung ca. (non-sm. cell) A549	4.0
CNS ca. (glio/astro) U-118-MG	4.1	Lung ca. (non-s.cell) NCI-H23	4.8
CNS ca. (astro) SW1783	2.0	Lung ca (non-s.cell) HOP-62	3.6
CNS ca.* (neuro; met) SK-N-AS	1.8	Lung ca. (non-s.cl) NCI-H522	8.1
CNS ca. (astro) SF-539	1.9	Lung ca. (squam.) SW 900	1.7
CNS ca. (astro) SNB-75	28.6	Lung ca. (squam.) NCI-H596	8.0
CNS ca. (glio) SNB-19	3.9	Mammary gland	0.8
CNS ca. (glio) U251	20.2	Breast ca.* (pl. effusion) MCF-7	1.1
CNS ca. (glio) SF-295	2.3	Breast ca.* (pl.ef) MDA-MB-231	2.9
Heart (fetal)	3.5	Breast ca.* (pl. effusion) T47D	2.7
Heart	3.9	Breast ca. BT-549	1.0
Fetal Skeletal	1.8	Breast ca. MDA-N	0.6
Skeletal muscle	0.7	Ovary	1.7
Bone marrow	0.4	Ovarian ca. OVCAR-3	2.4
Thymus	1.2	Ovarian ca. OVCAR-4	3.6
Spleen	0.1	Ovarian ca. OVCAR-5	3.7
Lymph node	0.8	Ovarian ca. OVCAR-8	1.7
Colorectal	3.6	Ovarian ca. IGROV-1	0.2
Stomach	0.5	Ovarian ca.* (ascites) SK-OV-3	7.5

Small intestine	0.4	Uterus	0.8
Colon ca. SW480	0.8	Placenta	0.2
Colon ca.* (SW480 met)SW620	4.7	Prostate	0.7
Colon ca. HT29	0.4	Prostate ca.* (bone met)PC-3	12.2
Colon ca. HCT-116	0.2	Testis	1.4
Colon ca. CaCo-2	1.7	Melanoma Hs688(A).T	9.3
83219 CC Well to Mod Diff (ODO3866)	7.0	Melanoma* (met) Hs688(B).T	**100.0**
Colon ca. HCC-2998	1.1	Melanoma UACC-62	0.8
Gastric ca.* (liver met) NCI-N87	1.3	Melanoma M14	0.5
Bladder	1.5	Melanoma LOX IMVI	1.9
Trachea	0.2	Melanoma* (met) SK-MEL-5	0.3
Kidney	3.8	Adipose	0.9

Table 25. Panel 2D

Tissue Name	Relative Expression(%) 2Dtm2408f_ag339	Relative Expression(%) 2dx4tm4720f_ag1391_a1
Normal Colon GENPAK 061003	4.3	20.9
83219 CC Well to Mod Diff (ODO3866)	6.3	18.6
83220 CC NAT (ODO3866)	3.4	5.4
83221 CC Gr.2 rectosigmoid (ODO3868)	0.9	2.3
83222 CC NAT (ODO3868)	3.5	4.5
83235 CC Mod Diff (ODO3920)	1.8	4.7
83236 CC NAT (ODO3920)	1.8	6.8
83237 CC Gr.2 ascend colon (ODO3921)	0.9	3.4
83238 CC NAT (ODO3921)	1.0	2.5
83241 CC from Partial Hepatectomy (ODO4309)	1.2	4.2
83242 Liver NAT (ODO4309)	1.4	1.0
87472 Colon mets to lung (OD04451-01)	4.4	3.5
87473 Lung NAT (OD04451-02)	0.1	3.3
Normal Prostate Clontech A+ 6546-1	6.0	26.3
84140 Prostate Cancer (OD04410)	6.7	11.4
84141 Prostate NAT (OD04410)	14.1	26.1
87073 Prostate Cancer (OD04720-01)	10.3	12.4
87074 Prostate NAT (OD04720-02)	16.6	21.3
Normal Lung GENPAK 061010	22.5	19.3
83239 Lung Met to Muscle (ODO4286)	2.0	1.8
83240 Muscle NAT (ODO4286)	7.3	2.5
84136 Lung Malignant Cancer (OD03126)	20.3	3.4
84137 Lung NAT (OD03126)	6.5	12.0

84871 Lung Cancer (OD04404)	3.5	4.2
84872 Lung NAT (OD04404)	2.6	5.6
84875 Lung Cancer (OD04565)	3.0	5.4
84876 Lung NAT (OD04565)	9.1	5.5
85950 Lung Cancer (OD04237-01)	11.8	10.3
85970 Lung NAT (OD04237-02)	2.1	6.9
83255 Ocular Mel Met to Liver (ODO4310)	7.2	6.3
83256 Liver NAT (ODO4310)	0.0	0.8
84139 Melanoma Mets to Lung (OD04321)	7.6	6.9
84138 Lung NAT (OD04321)	11.3	5.6
Normal Kidney GENPAK 061008	18.4	25.8
83786 Kidney Ca, Nuclear grade 2 (OD04338)	11.3	12.6
83787 Kidney NAT (OD04338)	6.3	14.9
83788 Kidney Ca Nuclear grade 1/2 (OD04339)	6.1	11.3
83789 Kidney NAT (OD04339)	15.5	16.6
83790 Kidney Ca, Clear cell type (OD04340)	14.4	11.5
83791 Kidney NAT (OD04340)	10.2	21.9
83792 Kidney Ca, Nuclear grade 3 (OD04348)	5.3	2.9
83793 Kidney NAT (OD04348)	14.0	11.8
87474 Kidney Cancer (OD04622-01)	5.0	2.0
87475 Kidney NAT (OD04622-03)	3.8	3.9
85973 Kidney Cancer (OD04450-01)	4.6	3.2
85974 Kidney NAT (OD04450-03)	6.1	11.9
Kidney Cancer Clontech 8120607	1.3	1.9
Kidney NAT Clontech 8120608	5.1	3.4
Kidney Cancer Clontech 8120613	3.6	1.2
Kidney NAT Clontech 8120614	2.6	7.6
Kidney Cancer Clontech 9010320	4.2	5.8
Kidney NAT Clontech 9010321	7.9	12.8
Normal Uterus GENPAK 061018	10.1	3.8
Uterus Cancer GENPAK 064011	21.0	11.8
Normal Thyroid Clontech A+ 6570-1	**100.0**	**100.0**
Thyroid Cancer GENPAK 064010	5.4	4.3
Thyroid Cancer INVITROGEN A302152	8.7	8.6
Thyroid NAT INVITROGEN A302153	57.8	55.4
Normal Breast GENPAK 061019	14.0	11.1
84877 Breast Cancer (OD04566)	9.9	6.1
85975 Breast Cancer (OD04590-01)	6.9	11.4
85976 Breast Cancer Mets (OD04590-03)	27.9	19.6
87070 Breast Cancer Metastasis (OD04655-05)	15.7	6.4
GENPAK Breast Cancer 064006	12.7	18.5
Breast Cancer Res. Gen. 1024	6.9	11.9

Breast Cancer Clontech 9100266	6.2	9.5
Breast NAT Clontech 9100265	6.0	5.7
Breast Cancer INVITROGEN A209073	8.3	9.1
Breast NAT INVITROGEN A2090734	2.2	6.0
Normal Liver GENPAK 061009	1.1	1.2
Liver Cancer GENPAK 064003	0.6	1.2
Liver Cancer Research Genetics RNA 1025	0.6	2.1
Liver Cancer Research Genetics RNA 1026	0.9	1.5
Paired Liver Cancer Tissue Research Genetics RNA 6004-T	0.5	1.5
Paired Liver Tissue Research Genetics RNA 6004-N	1.4	3.7
Paired Liver Cancer Tissue Research Genetics RNA 6005-T	1.0	0.8
Paired Liver Tissue Research Genetics RNA 6005-N	0.5	0.5
Normal Bladder GENPAK 061001	8.5	6.8
Bladder Cancer Research Genetics RNA 1023	17.2	27.6
Bladder Cancer INVITROGEN A302173	2.9	9.4
87071 Bladder Cancer (OD04718-01)	4.5	3.0
87072 Bladder Normal Adjacent (OD04718-03)	15.3	41.9
Normal Ovary Res. Gen.	1.1	5.0
Ovarian Cancer GENPAK 064008	3.2	10.9
87492 Ovary Cancer (OD04768-07)	7.6	10.7
87493 Ovary NAT (OD04768-08)	1.8	1.7
Normal Stomach GENPAK 061017	3.1	10.8
Gastric Cancer Clontech 9060358	0.8	3.4
NAT Stomach Clontech 9060359	2.0	2.4
Gastric Cancer Clontech 9060395	2.9	2.7
NAT Stomach Clontech 9060394	2.0	3.9
Gastric Cancer Clontech 9060397	3.5	6.3
NAT Stomach Clontech 9060396	0.7	1.5
Gastric Cancer GENPAK 064005	1.9	10.0

Table 26. Panel 4D

Tissue Name	Relative Expression(%) 4dx4tm5044f_ag2890_b1	Relative Expression(%) 4dtm2440f_ag1391	Relative Expression(%) 4Dtm2506f_ag1391
93768_Secondary Th1_anti-CD28/anti-CD3	1.2	1.9	4.5
93769_Secondary Th2_anti-CD28/anti-CD3	2.1	3.0	3.9
93770_Secondary Tr1_anti-CD28/anti-CD3	2.0	2.1	3.5
93573_Secondary Th1_resting day 4-6 in IL-2	0.4	0.3	0.6
93572_Secondary Th2_resting day 4-6 in IL-2	0.8	0.7	0.8
93571_Secondary Tr1_resting day 4-6 in IL-2	0.5	0.8	3.8
93568_primary Th1_anti-CD28/anti-CD3	1.0	2.1	2.3

93569_primary Th2_anti-CD28/anti-CD3	1.7	2.1	3.4
93570_primary Tr1_anti-CD28/anti-CD3	1.8	4.0	2.1
93565_primary Th1_resting dy 4-6 in IL-2	4.1	2.9	4.4
93566_primary Th2_resting dy 4-6 in IL-2	2.0	2.1	2.8
93567_primary Tr1_resting dy 4-6 in IL-2	2.6	2.8	2.3
93351_CD45RA CD4 lymphocyte_anti-CD28/anti-CD3	16.5	21.8	25.5
93352_CD45RO CD4 lymphocyte_anti-CD28/anti-CD3	1.3	2.8	3.0
93251_CD8 Lymphocytes_anti-CD28/anti-CD3	1.7	1.1	2.0
93353_chronic CD8 Lymphocytes 2ry_resting dy 4-6 in IL-2	1.6	1.9	2.5
93574_chronic CD8 Lymphocytes 2ry_activated CD3/CD28	1.2	1.6	0.8
93354_CD4_none	0.8	0.8	0.6
93252_Secondary Th1/Th2/Tr1_anti-CD95 CH11	1.7	3.1	2.0
93103_LAK cells_resting	1.5	1.9	2.6
93788_LAK cells_IL-2	2.4	4.7	3.0
93787_LAK cells_IL-2+IL-12	2.1	2.6	1.6
93789_LAK cells_IL-2+IFN gamma	3.9	3.2	7.0
93790_LAK cells_IL-2+ IL-18	3.3	2.2	4.3
93104_LAK cells_PMA/ionomycin and IL-18	1.3	1.0	1.0
93578_NK Cells IL-2_resting	2.5	1.9	3.5
93109_Mixed Lymphocyte Reaction_Two Way MLR	2.9	1.4	2.2
93110_Mixed Lymphocyte Reaction_Two Way MLR	1.8	1.1	1.8
93111_Mixed Lymphocyte Reaction_Two Way MLR	1.4	1.3	1.1
93112_Mononuclear Cells (PBMCs)_resting	0.2	0.3	1.1
93113_Mononuclear Cells (PBMCs)_PWM	6.7	5.6	7.4
93114_Mononuclear Cells (PBMCs)_PHA-L	1.8	3.3	3.3
93249_Ramos (B cell)_none	2.9	5.9	3.4
93250_Ramos (B cell)_ionomycin	7.8	8.8	9.0
93349_B lymphocytes_PWM	3.4	3.1	4.4
93350_B lymphoytes_CD40L and IL-4	3.7	3.4	3.4
92665_EOL-1 (Eosinophil)_dbcAMP differentiated	3.5	2.6	5.7
93248_EOL-1 (Eosinophil)_dbcAMP/PMAionomycin	5.1	5.6	7.3
93356_Dendritic Cells_none	0.8	1.0	1.0
93355_Dendritic Cells_LPS 100 ng/ml	1.2	1.4	1.6
93775_Dendritic Cells_anti-CD40	0.8	0.9	1.4
93774_Monocytes_resting	2.0	1.0	1.7

93776_Monocytes_LPS 50 ng/ml	2.5	4.1	7.0
93581_Macrophages_resting	1.3	1.7	1.8
93582_Macrophages_LPS 100 ng/ml	0.8	3.0	1.5
93098_HUVEC (Endothelial)_none	3.1	2.5	3.9
93099_HUVEC (Endothelial)_starved	4.8	4.8	10.5
93100_HUVEC (Endothelial)_IL-1b	1.6	2.1	2.4
93779_HUVEC (Endothelial)_IFN gamma	3.7	4.3	6.2
93102_HUVEC (Endothelial)_TNF alpha + IFN gamma	1.9	1.1	2.4
93101_HUVEC (Endothelial)_TNF alpha + IL4	2.7	1.7	3.1
93781_HUVEC (Endothelial)_IL-11	3.9	3.8	3.3
93583_Lung Microvascular Endothelial Cells_none	7.3	6.0	17.3
93584_Lung Microvascular Endothelial Cells_TNFa (4 ng/ml) and IL1b (1 ng/ml)	6.3	5.0	8.8
92662_Microvascular Dermal endothelium_none	6.8	11.8	12.9
92663_Microsvasular Dermal endothelium_TNFa (4 ng/ml) and IL1b (1 ng/ml)	3.9	4.2	3.9
93773_Bronchial epithelium_TNFa (4 ng/ml) and IL1b (1 ng/ml) **	1.3	1.0	1.2
93347_Small Airway Epithelium_none	0.9	0.1	0.5
93348_Small Airway Epithelium_TNFa (4 ng/ml) and IL1b (1 ng/ml)	2.8	2.9	4.0
92668_Coronery Artery SMC_resting	3.4	4.5	5.5
92669_Coronery Artery SMC_TNFa (4 ng/ml) and IL1b (1 ng/ml)	1.3	2.3	1.5
93107_astrocytes_resting	6.0	4.8	8.2
93108_astrocytes_TNFa (4 ng/ml) and IL1b (1 ng/ml)	10.6	10.0	12.6
92666_KU-812 (Basophil)_resting	2.3	2.5	3.8
92667_KU-812 (Basophil)_PMA/ionoycin	6.3	5.4	10.7
93579_CCD1106 (Keratinocytes)_none	2.5	1.0	3.1
93580_CCD1106 (Keratinocytes)_TNFa and IFNg **	0.6	1.3	0.7
93791_Liver Cirrhosis	0.9	1.1	1.7
93792_Lupus Kidney	2.7	9.0	7.5
93577_NCI-H292	3.7	3.7	4.7
93358_NCI-H292_IL-4	6.4	5.8	11.7
93360_NCI-H292_IL-9	5.4	5.3	9.4
93359_NCI-H292_IL-13	3.5	4.1	5.1
93357_NCI-H292_IFN gamma	3.5	2.5	4.7
93777_HPAEC_-	1.6	3.4	2.9
93778_HPAEC_IL-1 beta/TNA alpha	2.2	2.8	4.0

93254_Normal Human Lung Fibroblast_none	2.3	2.6	2.9
93253_Normal Human Lung Fibroblast_TNFa (4 ng/ml) and IL-1b (1 ng/ml)	2.0	1.4	1.0
93257_Normal Human Lung Fibroblast_IL-4	13.0	10.6	17.0
93256_Normal Human Lung Fibroblast_IL-9	4.1	2.2	4.3
93255_Normal Human Lung Fibroblast_IL-13	9.9	17.3	10.1
93258_Normal Human Lung Fibroblast_IFN gamma	3.8	3.5	4.0
93106_Dermal Fibroblasts CCD1070_resting	**100.0**	**100.0**	**100.0**
93361_Dermal Fibroblasts CCD1070_TNF alpha 4 ng/ml	94.1	93.3	91.4
93105_Dermal Fibroblasts CCD1070_IL-1 beta 1 ng/ml	35.5	41.5	33.9
93772_dermal fibroblast_IFN gamma	4.7	3.3	4.0
93771_dermal fibroblast_IL-4	13.9	18.3	20.2
93259_IBD Colitis 1**	1.2	1.4	0.5
93260_IBD Colitis 2	0.3	0.7	0.5
93261_IBD Crohns	0.3	0.3	0.4
735010_Colon_normal	1.8	3.0	4.2
735019_Lung_none	2.4	1.8	3.3
64028-1_Thymus_none	13.0	24.1	16.8
64030-1_Kidney_none	11.4	10.7	13.0

Table 27. Table 4.1D

Tissue Name	Relative Expression(%) 4.1dx4tm6545f_ag339_a1	Tissue Name	Relative Expression(%) 4.1dx4tm6545f_ag339_a1
93768_Secondary Th1_anti-CD28/anti-CD3	7.5	93100_HUVEC (Endothelial)_IL-1b	5.8
93769_Secondary Th2_anti-CD28/anti-CD3	7.3	93779_HUVEC (Endothelial)_IFN gamma	7.7
93770_Secondary Tr1_anti-CD28/anti-CD3	23.0	93102_HUVEC (Endothelial)_TNF alpha + IFN gamma	8.7
93573_Secondary Th1_resting day 4-6 in IL-2	3.1	93101_HUVEC (Endothelial)_TNF alpha + IL4	5.7
93572_Secondary Th2_resting day 4-6 in IL-2	2.4	93781_HUVEC (Endothelial)_IL-11	13.1
93571_Secondary Tr1_resting day 4-6 in IL-2	1.2	93583_Lung Microvascular Endothelial Cells_none	26.3
93568_primary Th1_anti-CD28/anti-CD3	3.3	93584_Lung Microvascular Endothelial Cells_TNFa (4 ng/ml) and IL1b (1 ng/ml)	17.6
93569_primary Th2_anti-CD28/anti-CD3	5.8	92662_Microvascular Dermal endothelium_none	14.0

93570_primary Tr1_anti-CD28/anti-CD3	6.3	92663_Microsvasular Dermal endothelium_TNFa (4 ng/ml) and IL1b (1 ng/ml)	7.2
93565_primary Th1_resting dy 4-6 in IL-2	0.5	93773_Bronchial epithelium_TNFa (4 ng/ml) and IL1b (1 ng/ml) **	3.2
93566_primary Th2_resting dy 4-6 in IL-2	2.5	93347_Small Airway Epithelium_none	1.3
93567_primary Tr1_resting dy 4-6 in IL-2	5.3	93348_Small Airway Epithelium_TNFa (4 ng/ml) and IL1b (1 ng/ml)	2.1
93351_CD45RA CD4 lymphocyte_anti-CD28/anti-CD3	45.3	92668_Coronery Artery SMC_resting	10.4
93352_CD45RO CD4 lymphocyte_anti-CD28/anti-CD3	6.3	92669_Coronery Artery SMC_TNFa (4 ng/ml) and IL1b (1 ng/ml)	4.7
93251_CD8 Lymphocytes_anti-CD28/anti-CD3	8.1	93107_astrocytes_resting	10.9
93353_chronic CD8 Lymphocytes 2ry_resting dy 4-6 in IL-2	10.4	93108_astrocytes_TNFa (4 ng/ml) and IL1b (1 ng/ml)	16.1
93574_chronic CD8 Lymphocytes 2ry_activated CD3/CD28	2.4	92666_KU-812 (Basophil)_resting	8.6
93354_CD4_none	3.7	92667_KU-812 (Basophil)_PMA/ionoycin	16.3
93252_Secondary Th1/Th2/Tr1_anti-CD95 CH11	6.8	93579_CCD1106 (Keratinocytes)_none	9.2
93103_LAK cells_resting	6.2	93580_CCD1106 (Keratinocytes)_TNFa and IFNg **	4.6
93788_LAK cells_IL-2	6.8	93791_Liver Cirrhosis	1.1
93787_LAK cells_IL-2+IL-12	2.5	93577_NCI-H292	11.5
93789_LAK cells_IL-2+IFN gamma	7.5	93358_NCI-H292_IL-4	8.6
93790_LAK cells_IL-2+ IL-18	3.9	93360_NCI-H292_IL-9	13.2
93104_LAK cells_PMA/ionomycin and IL-18	3.2	93359_NCI-H292_IL-13	16.6
93578_NK Cells IL-2_resting	12.2	93357_NCI-H292_IFN gamma	14.4
93109_Mixed Lymphocyte Reaction_Two Way MLR	6.6	93777_HPAEC_-	7.4
93110_Mixed Lymphocyte Reaction_Two Way MLR	4.3	93778_HPAEC_IL-1 beta/TNA alpha	16.0
93111_Mixed Lymphocyte Reaction_Two Way MLR	4.9	93254_Normal Human Lung Fibroblast_none	4.8
93112_Mononuclear Cells	0.5	93253_Normal Human Lung	4.1

(PBMCs)_resting		Fibroblast_TNFa (4 ng/ml) and IL-1b (1 ng/ml)	
93113_Mononuclear Cells (PBMCs)_PWM	7.4	93257_Normal Human Lung Fibroblast_IL-4	14.4
93114_Mononuclear Cells (PBMCs)_PHA-L	5.5	93256_Normal Human Lung Fibroblast_IL-9	7.6
93249_Ramos (B cell)_none	10.3	93255_Normal Human Lung Fibroblast_IL-13	11.4
93250_Ramos (B cell)_ionomycin	14.4	93258_Normal Human Lung Fibroblast_IFN gamma	7.5
93349_B lymphocytes_PWM	3.8	93106_Dermal Fibroblasts CCD1070_resting	**100.0**
93350_B lymphoytes_CD40L and IL-4	8.4	93361_Dermal Fibroblasts CCD1070_TNF alpha 4 ng/ml	85.2
92665_EOL-1 (Eosinophil)_dbcAMP differentiated	20.0	93105_Dermal Fibroblasts CCD1070_IL-1 beta 1 ng/ml	32.5
93248_EOL-1 (Eosinophil)_dbcAMP/PMAionomycin	15.3	93772_dermal fibroblast_IFN gamma	11.7
93356_Dendritic Cells_none	3.1	93771_dermal fibroblast_IL-4	24.8
93355_Dendritic Cells_LPS 100 ng/ml	4.9	93892_Dermal fibroblasts_none	8.5
93775_Dendritic Cells_anti-CD40	8.3	99202_Neutrophils_TNFa+LPS	6.4
93774_Monocytes_resting	2.1	99203_Neutrophils_none	3.5
93776_Monocytes_LPS 50 ng/ml	21.5	735010_Colon_normal	3.1
93581_Macrophages_resting	1.8	735019_Lung_none	9.5
93582_Macrophages_LPS 100 ng/ml	4.6	64028-1_Thymus_none	17.4
93098_HUVEC (Endothelial)_none	3.8	64030-1_Kidney_none	59.4
93099_HUVEC (Endothelial)_starved	8.5		

Summary of Panels

Probes Ag192/Ag339 in Panel 1 show distinct differences in replicate runs. However, expression of the GM-Z97832B gene appears to be distributed through all the samples. There appears to be strong GM-Z97832B gene expression in both a metastatic melanoma cell line, when compared to a cell line derived from a primary melanoma, as well as a metastatic prostate cancer cell line. This suggests that this gene might be associated with the process of metastasis and, as such, therapeutic modulation of its expression may be useful in the treatment of metastatic cancer. The GM-Z97832B gene is also well expressed in

hypothalamus, pancreas, skeletal muscle, liver and pituitary. Because these are metabolic and/or insulin-responsive tissues, this gene may be an important drug target for the treatment or prevention of metabolic diseases like diabetes and obesity. In addition, since appetite and satiety centers are localized to the hypothalamus and pituitary gland, this gene also may be involved in appetite regulation.

Probe Ag192 in Panel 1.1 evidences widespread expression of the GM-Z97832B gene. Specifically, this gene seems to be most highly expressed in samples derived from thyroid, an astrocytoma cell line and a prostate cancer cell line. Thus, therapeutic modulation of this gene may be useful in treating prostate cancer, brain cancer or diseases of the thyroid. These results are consistent with those from Panel 1 as are the expressions of the GM-Z97832B gene in hypothalamus, pancreas, skeletal muscle, liver and pituitary samples.

Probe Ag2890 in Panel 1.3D shows that the GM-Z97832B transcript appears to be present in the most samples. Expression of this gene is highest in a metastatic melanoma cell line when compared to primary melanoma cell lines derived from the same patient. Thus, therapeutic modulation of this gene may prove useful in the treatment of metastatic melanoma. In addition, high expression is detected in thyroid samples, suggesting a possible role for the GM-Z97832B gene in thyroid function and disorders. Moderate GM-Z97832B gene expression in brain samples may suggest a role in neurological function and diseases.

Probe Ag339/Ag1391 indicates that expression of the GM-Z97832B gene is found in a wide variety of samples in Panel 2D. Of particular interest is the observation that this gene is highly expressed in normal thyroid tissues when compared to thyroid cancers. Consequently, this gene may play a role in normal thyroid homeostasis, and its modulation may be of use in the treatment of thyroid cancer.

Probes Ag2890/Ag1391 and Ag339 in Panels 4D and 4.1D respectively show that the GM-Z97832B gene is expressed in dermal fibroblasts and is inhibited by treatment with IL-4 or gamma interferon. The transcript is induced in newly activated naïve T cells (CD45RA), but not in treated or fully polarized T cells. The notch/serrate molecules are involved in cellular fate determination (see Development, 2000 Mar; 127(6):1291-302 in which it is shown that the Abruptex domain of Notch regulates negative interactions between Notch, its ligands and Fringe), and may be important in initiating the sequence of events that leads to differentiation and/or activation. Antagonistic antibody or small molecule therapeutics can

block T cell and dermal fibroblast activation/differentiation that result from specific activation, i.e., T cell activation through the T cell receptor via CD3/CD28 or cytokine activation alone (IL-4/gamma interferon). Such blocking action holds important implications for the treatment of psoriasis, asthma, arthritis and other autoimmune diseases.

EXAMPLE 6. EXPRESSION OF A POLY11 NUCLEIC ACID IN CELLS AND TISSUES

The quantitative expression of various clones was assessed in 41 normal and 55 tumor samples (in most cases, the samples are those identified in Table BB) by real time quantitative PCR (TAQMAN®) performed on a Perkin-Elmer Biosystems ABI PRISM® 7700 Sequence Detection System. The following abbreviations are used:

ca. = carcinoma,
* = established from metastasis,
met = metastasis,
s cell var= small cell variant,
non-s = non-sm =non-small,
squam = squamous,
pl. eff = pl effusion = pleural effusion,
glio = glioma,
astro = astrocytoma, and
neuro = neuroblastoma.

First, 96 RNA samples were normalized to ß-actin and GAPDH. RNA (~50 ng total or ~1 ng polyA+) was converted to cDNA using the TAQMAN® Reverse Transcription Reagents Kit (PE Biosystems, Foster City, CA; cat # N808-0234) and random hexamers according to the manufacturer's protocol. Reactions were performed in 20 ul and incubated for 30 min. at 48^{0}C. cDNA (5 ul) was then transferred to a separate plate for the TAQMAN® reaction using b-actin and GAPDH TAQMAN® Assay Reagents (PE Biosystems; cat. #'s 4310881E and 4310884E, respectively) and TAQMAN® universal PCR Master Mix (PE Biosystems; cat # 4304447) according to the manufacturer's protocol. Reactions were performed in 25 ul using the following parameters: 2 min. at 50^{0}C; 10 min. at 95^{0}C; 15 sec. at 95^{0}C/1 min. at 60^{0}C (40 cycles). Results were recorded as CT values (cycle at which a given sample crosses a threshold level of fluorescence) using a log scale, with the difference in RNA concentration

between a given sample and the sample with the lowest CT value being represented as 2 to the power of delta CT. The percent relative expression is then obtained by taking the reciprocal of this RNA difference and multiplying by 100. The average CT values obtained for ß-actin and GAPDH were used to normalize RNA samples. The RNA sample generating the highest CT value required no further diluting, while all other samples were diluted relative to this sample according to their b-actin /GAPDH average CT values.

Normalized RNA (5 ul) was converted to cDNA and analyzed via TAQMAN® using One Step RT-PCR Master Mix Reagents (PE Biosystems; cat. # 4309169) and gene-specific primers according to the manufacturer's instructions. Probes and primers were designed for each assay according to Perkin Elmer Biosystem's *Primer Express* Software package (version I for Apple Computer's Macintosh Power PC) using the sequence of clone 10326230.0.38 as input. Default settings were used for reaction conditions and the following parameters were set before selecting primers: primer concentration = 250 nM, primer melting temperature (T_m) range = 58°-60° C, primer optimal Tm = 59° C, maximum primer difference = 2° C, probe does not have 5' G, probe T_m must be 10° C greater than primer T_m, amplicon size 75 bp to 100 bp. The primers and probe selected were:

Ag 373 (F): 5'-GTGTGTTCCTCTCGACTGTGGA-3' (SEQ ID NO:47)
Ag 373 (R): 5'-GACCCTTGGACCCTACTTCAAA-3' (SEQ ID NO:48)
Ag 373 (P): TET-5'-CCCCGATCCAGAATGGCTTCATGA-3'-TAMRA. (SEQ ID NO:49)

They were synthesized by Synthegen (Houston, TX, USA). Probes were double purified by HPLC to remove uncoupled dye and evaluated by mass spectroscopy to verify coupling of reporter and quencher dyes to the 5' and 3' ends of the probe, respectively. Their final concentrations were: forward and reverse primers, 900 nM each, and probe, 200nM.

PCR conditions: Normalized RNA from each tissue and each cell line was spotted in each well of a 96 well PCR plate (Perkin Elmer Biosystems). PCR cocktails including two probes (SEQX-specific and another gene-specific probe multiplexed with the SEQX probe) were set up using 1X TaqMan™ PCR Master Mix for the PE Biosystems 7700, with 5 mM MgCl2, dNTPs (dA, G, C, U at 1:1:1:2 ratios), 0.25 U/ml AmpliTaq Gold™ (PE Biosystems), and 0.4 U/ml RNase inhibitor, and 0.25 U/ml reverse transcriptase. Reverse transcription was performed at 48° C for 30 minutes followed by amplification/PCR cycles as follows: 95° C 10

min, then 40 cycles of 95° C for 15 seconds, 60° C for 1 minute.

The results are shown in Table 19.

Table 19. Real Time Expression Analysis of 10327789_1.

Number	Type of Tissue	Relative Expression (%)
1	Endothelial cells	0.00
2	Endothelial cells (treated)	0.00
3	Pancreas	0.02
4	Pancreatic ca. CAPAN 2	0.00
5	Adipose	67.83
6	Adrenal gland	1.94
7	Thyroid	0.05
8	Salavary gland	0.00
9	Pituitary gland	0.81
10	Brain (fetal)	0.00
11	Brain (whole)	0.55
12	Brain (amygdala)	0.00
13	Brain (cerebellum)	18.82
14	Brain (hippocampus)	0.00
15	Brain (hypothalamus)	0.00
16	Brain (substantia nigra)	0.05
17	Brain (thalamus)	0.00
18	Spinal cord	0.00
19	CNS ca. (glio/astro) U87-MG	0.07
20	CNS ca. (glio/astro) U-118-MG	0.02
21	CNS ca. (astro) SW1783	0.00
22	CNS ca.* (neuro; met) SK-N-AS	0.00
23	CNS ca. (astro) SF-539	0.13
24	CNS ca. (astro) SNB-75	0.00
25	CNS ca. (glio) SNB-19	0.00
26	CNS ca. (glio) U251	0.00
27	CNS ca. (glio) SF-295	0.00
28	Heart	0.49
29	Skeletal muscle	0.00
30	Bone marrow	0.00
31	Thymus	0.98
32	Spleen	0.43
33	Lymph node	3.90
34	Colon (ascending)	1.07
35	Stomach	2.76
36	Small intestine	2.35
37	Colon ca. SW480	0.00
38	Colon ca.* (SW480 met) SW620	0.00
39	Colon ca. HT29	0.00
40	Colon ca. HCT-116	0.00
41	Colon ca. CaCo-2	0.24
42	Colon ca. HCT-15	0.00

43	Colon ca. HCC-2998	0.00
44	Gastric ca.* (liver met) NCI-N87	0.00
45	Bladder	0.00
46	Trachea	0.97
47	Kidney	0.15
48	Kidney (fetal)	0.40
49	Renal ca. 786-0	0.00
50	Renal ca. A498	0.00
51	Renal ca. RXF 393	0.00
52	Renal ca. ACHN	0.00
53	Renal ca. UO-31	0.00
54	Renal ca. TK-10	0.00
55	Liver	0.00
56	Liver (fetal)	0.00
57	Liver ca. (hepatoblast) HepG2	0.00
58	Lung	14.76
59	Lung (fetal)	0.50
60	Lung ca. (small cell) LX-1	0.00
61	Lung ca. (small cell) NCI-H69	0.00
62	Lung ca. (s.cell var.) SHP-77	0.00
63	Lung ca. (large cell)NCI-H460	0.00
64	Lung ca. (non-sm. cell) A549	0.01
65	Lung ca. (non-s.cell) NCI-H23	0.25
66	Lung ca (non-s.cell) HOP-62	0.00
67	Lung ca. (non-s.cl) NCI-H522	0.00
68	Lung ca. (squam.) SW 900	0.00
69	Lung ca. (squam.) NCI-H596	0.00
70	Mammary gland	20.45
71	Breast ca.* (pl. effusion) MCF-7	0.00
72	Breast ca.* (pl.ef) MDA-MB-231	0.00
73	Breast ca.* (pl. effusion) T47D	0.00
74	Breast ca. BT-549	0.00
75	Breast ca. MDA-N	0.00
76	Ovary	1.94
77	Ovarian ca. OVCAR-3	0.00
78	Ovarian ca. OVCAR-4	0.00
79	Ovarian ca. OVCAR-5	0.00
80	Ovarian ca. OVCAR-8	0.00
81	Ovarian ca. IGROV-1	0.00
82	Ovarian ca.* (ascites) SK-OV-3	0.00
83	Myometrium	8.96
84	Uterus	0.58
85	Placenta	100.00
86	Prostate	0.00
87	Prostate ca.* (bone met)PC-3	0.00
88	Testis	1.34
89	Melanoma Hs688(A).T	0.03
90	Melanoma* (met) Hs688(B).T	0.36
91	Melanoma UACC-62	0.00
92	Melanoma M14	0.00

93	Melanoma	LOX IMVI	0.00
94	Melanoma* (met)	SK-MEL-5	0.00
95	Melanoma	SK-MEL-28	0.00
96	Melanoma	UACC-257	0.00

EXAMPLE 7. IDENTIFICATION OF POLY15 AND POLY16.

For POLY15, PCR primers were designed by starting at the most upstream sequence available, for the forward primer, and at the most downstream sequence available for the reverse primer. In each case, the sequence was examined, walking inward from the respective termini toward the coding sequence, until a suitable sequence that is either unique or highly selective was encountered, or, in the case of the reverse primer, until the stop codon was reached. Such suitable sequences were then employed as the forward and reverse primers in a PCR amplification based on a library containing a wide range of cDNA species. The resulting amplicon was gel purified, cloned and sequenced to high redundancy to provide the sequence reported below, which is designated POLY15 (Accession Number hnh0778p17_A1.) POLY15 exhibits no change at the ORF level with respect to h_nh0778p17_A. A physical clone, clone hnh0778p17_A.699002.A7, was identified that covers the entire ORF.

For POLY16, PCR primers were designed by starting at the most upstream sequence available, for the forward primer, and at the most downstream sequence available for the reverse primer. In each case, the sequence was examined, walking inward from the respective termini toward the coding sequence, until a suitable sequence that is either unique or highly selective was encountered, or, in the case of the reverse primer, until the stop codon was reached. Such suitable sequences were then employed as the forward and reverse primers in a PCR amplification based on a library containing a wide range of cDNA species. The resulting amplicon was gel purified, cloned and sequenced to high redundancy to provide the sequence reported below, which is designated Accession Number hnh0778p17_A1. Clone hnh0778p17_A1 exhibits no change at the ORF level with respect to h_nh0778p17_A. A physical clone, clone hnh0778p17_A.699002.A7, was identified that covers the entire ORF. This procedure also yielded POLY16 (CG55655-02) that is 100% identical to h_nh0778p17_A.

Other Embodiments

While the invention has been described in conjunction with the detailed description thereof, the foregoing description is intended to illustrate and not limit the scope of the invention, which is defined by the scope of the appended claims. Other aspects, advantages, and modifications are within the scope of the following claims.